STATISTICAL IMAGE PROCESSING:
RESTORATION AND RECONSTRUCTION

by

Thomas James Hebert

A Dissertation Presented to the

FACULTY OF THE GRADUATE SCHOOL

UNIVERSITY OF SOUTHERN CALIFORNIA

In Partial Fulfillment of the

Requirements for the Degree

DOCTOR OF PHILOSOPHY

(Electrical Engineering)

May 1989

UNIVERSITY OF SOUTHERN CALIFORNIA
THE GRADUATE SCHOOL
UNIVERSITY PARK
LOS ANGELES, CALIFORNIA 90089

This dissertation, written by

Thomas James Hebert

under the direction of his..... Dissertation Committee, and approved by all its members, has been presented to and accepted by The Graduate School, in partial fulfillment of requirements for the degree of

DOCTOR OF PHILOSOPHY

Barbara Schroeder

Dean of Graduate Studies

Date April 19, 1989

DISSERTATION COMMITTEE

Richard Luh

Chairperson

Ramon Challenger

Cecil Snyder

USC-SIPI REPORT # 145

Statistical Image Processing: Restoration and Reconstruction

By

Thomas James Hebert

Signal and Image Processing Institute
UNIVERSITY OF SOUTHERN CALIFORNIA
Department of Electrical Engineering-Systems
Powell Hall of Engineering
University Park/MC-0272
Los Angeles, CA 90089 U.S.A.

April 1989

Abstract

This dissertation develops statistical approaches to image restoration and reconstruction. A statistic-based stopping criterion for iterative least squares and maximum likelihood image reconstruction is developed to regularize the solution by effecting a trade-off between the maximum likelihood or least squares solution and a maximally smooth uniform image. Bayesian estimation is embraced as a probabilistic method of incorporating qualitative information regarding the local smoothness of images. Markov random field priors in the form of Gibbs distributions are formulated to incorporate local correlations and smoothness yet allow abrupt changes to occur across a boundary between two regions in an image. A computationally efficient algorithm for Bayesian image reconstruction based on Gibbs distribution priors is derived. This algorithm combines the complete/incomplete data formulation of the expectation - maximization approach, coordinate ascent, the optimal step-size of a maximum likelihood algorithm derived along parallel lines, and a method of adjusting the step size if necessary. A statistic-based approach to selection of the Gibbs prior parameter is developed and its potential for selecting the optimal parameter with respect to L_1 and L_2 restoration error is evaluated. The resulting solution is set in a control systems framework and a feedback algorithm for use in conjunction with any deterministic MAP algorithm for simultaneous parameter selection and image restoration or reconstruction is presented. Some problems in emission tomography are also addressed. An intermediate polar pixel representation for the unknown 3-D image is used to reduce the computational requirements of iterative reconstruction algorithms. An efficient method of directly including attenuation in iterative reconstruction algorithms is suggested. An experimental emission imaging system with a spatially varying point source response is examined. A probabilistically based factorization of the system matrix for the

experimental system is introduced allowing the computational demands of a class of iterative algorithms to be reduced by several orders of magnitude. A modification to a well known maximum likelihood algorithm is introduced to increase the rate of convergence of the likelihood by a factor of five.

To my better half.

Acknowledgements

I would first like to thank Dr. Jerry Mendel for nominating me for the American Electronics Association fellowship in 1984. This four year funding provided the means for me to attend the University of Southern California. The funds for this fellowship were donated by Northrop Corporation. I would like to thank my Ph.D. advisor Dr. Richard Leahy with whom I worked for 43 short months. I would also like to thank Dr. Leahy, Dr. Manbir Singh of the Radiology Department, and the National Institute of Health for the funding I received as a research assistant from September '86 through December '88.

The Signal and Image Processing Institute at USC has been a wonderful working environment. This is due largely to Dr. Sandy Sawchuk, Dr. Rama Chellapa, and Dr. Allan Weber. All of the graduate students at SIPI have benefited, proving that the trickle-down theory does occasionally work. Perhaps most importantly, I would like to thank all of the hard-working EE graduate students at SIPI who contributed to a colorful work environment. I am proud to call them my friends. Of these, one could not ask for a better officemate than Xiao Hong Yan. I thank her for that and wish her much luck.

Finally, the lesson I have carried with me throughout my graduate work is that perseverance, perspiration, professionalism, and principles payoff.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation and Goals

Many estimation problems are ill-posed in the sense defined by Hadamard (1923 Lectures on Cauchy's Problem, Yale Univ. Press, New Haven Conn.). A problem is said to be well-posed if its solution exists, is unique, and depends continuously on the data in a manner such that the solution is not sensitive to small perturbations of the data. With an ill-posed problem, a classical least squares solution may be unacceptably sensitive to noise or a maximum likelihood solution may not be unique.

A variety of approaches have been developed over the years to reformulate ill-posed estimation problems as well-posed ones. These generally take the form of imposing additional constraints on the solution, Tikhonov regularization, or incorporating prior information through Bayesian estimation. These approaches have been applied in many different areas of engineering, mathematics, and physics. The motivation behind the development of these approaches may be basically alike while the criterion to resolve the problem leads to differing solutions.

An alternative motivation exists for the application of Bayesian estimation or regularization techniques. That is, an investigator may have some insight into the proper structure of the solution or some quality which the solution should possess. In this case, it is desirable to include this information, along with the data, in the estimation process. Regularization achieves this through a regularizing function which is intended to impose the quality or structure on the solution to a degree specified by the regularization parameter. Bayesian approaches achieve this by

specification of a prior distribution for the solution. The prior distribution function should place increased probability on solutions which possess the known structure or quality, and decreased probability on solutions that do not. The prior influences the solution through the posterior distribution as defined by Bayes rule.

Image restoration and image reconstruction deal with the estimation of an undegraded image from a set of measurements of the image which have been degraded by one or more factors such as noise or blur. In image restoration, the set of observations are simply the noisy, degraded image itself. The dimension of the observation set is therefore the same as the dimension of the image. In image reconstruction, the observations are a set of measurements which are **indirectly** related to the image in a specified manner. The dimension of the data may differ dramatically from that of the undegraded image.

Since image restoration and reconstruction problems are typically ill-posed, it is not suprising that Bayesian estimation and regularization techniques are rapidly becoming very popular in image processing. Our growing insight into the fundamental structure of images reveals that images can be well modeled as consisting of many segmented regions with a degree of smoothness within each region. In addition, images generally possess only local correlations. In response to these insights, a class of probabilistic image models which consider only local dependencies in the image have been developed. These Markov random field models were developed as an extension of a Markov process to multi-dimensions. Markov random field models are currently being examined for their usefulness in Bayesian image restoration, reconstruction, and segmentation.

The application of a Bayesian or regularization approach requires a choice of some key parameters. In regularization, a regularizing parameter needs to be

chosen. In Bayesian estimation a prior distribution and its parameters need to be chosen. The relative success of these approaches may be dependent upon a favorable selection of these parameters. In general, the same parameter values will not be suitable for all data sets. The strategy of selecting appropriate values for these parameters by trial and error introduces a subjective component to the estimation process. If we examine certain applications in medical signal processing, we find that a doctors diagnosis may vary with the selection of a key parameter in a particular signal processing application. A suitable choice of a given parameter may lead to a correct diagnosis while an unsuitable choice may lead to a missed or incorrect diagnosis. It is therefore important to specify methods for selection of all of the parameters which are instrumental in the estimation process. Furthur, it is intended that these methods of parameter selection be more reliable than selection based on the skill and experience of a particular investigator.

In light of the above discussion, the work contained in this thesis has been directed towards the following three goals:

(1) To develop priors which are meaningful for image restoration and reconstruction and are easily incorporated into Bayesian estimation.

(2) To develop robust statistical methods for the selection of all prior parameters.

(3) To develop computationally efficient algorithms for Bayesian reconstruction.

1.2 Overview

This dissertation is divided into eight chapters. The first chapter contains a statement of and a motivation for the goals of this thesis. This is followed by an overview of the dissertation, and a statement of the contributions made by this work. The second chapter presents the observation models which are addressed as

well as several of the many applications which fit within these observation models. The third chapter examines probabilistic approaches to image restoration and reconstruction. This includes maximum likelihood and Bayesian reconstruction, Markov random fields and their description by Gibbs distributions, as well as established algorithms for their solution. Chapters four through seven present the original work by this author. The fourth chapter presents a statistic based stopping criterion for iterative reconstruction algorithms. The motivating factor in this approach is to regularize the solution by effecting a trade-off between the maximum likelihood solution and a maximally smooth uniform image. In the fifth chapter an algorithm for Bayesian restoration and reconstruction based on an complete/incomplete data formulation is developed. The sixth chapter contains a statistical method for the selection of the parameter of Gibbs functions in Bayesian restoration and reconstruction. This method requires the updating of the parameter as the iterative restoration or reconstruction proceeds. A simple algorithm for implementing this method is included in the chapter. The seventh chapter addresses some concerns in emission tomography. It contains an approach by which we have been able to decrease the computation required by popular iterative reconstruction algorithms in the field of emission tomography. Attenuation effects are addressed and an efficient method of incorporating attenuation coefficients in iterative reconstruction algorithms is suggested. Both simulated and experimental reconstructions are used to evaluate these approaches. This chapter also presents a step-by-step approach to developing a stochastic model for a prototype imaging system now under development at this university. Through a matrix factorization based upon a probabilistic model of the imaging modality, the computation required to implement a popular maximum likelihood algorithm is decreased by several orders of magnitude. A modification to a ML algorithm which increases the rate of convergence of the likelihood by a

factor of five is introduced. Chapter eight draws conclusions on the work presented and offers areas for furthur research.

1.3 Contributions

The contributions of this author to image restoration and reconstruction consist of the following:

(1) In image restoration and image reconstruction, maximum likelihood estimation is often an ill-conditioned problem so that the solution image possesses an excessive, non-smooth quality. Iterative restoration and reconstruction using an iteration stopping criterion to achieve a regularization between the maximum likelihood solution and a uniformly smooth image is presented. This approach imposes a degree of smoothness on the solution and results in a reduction in the required computation. Through Monte Carlo simulations an improvement over the maximum likelihood solution in terms of both absolute error and squared error is demonstrated.

(2) When working with three dimensional images or sequences of two dimensional images, the computational demand of simulated annealing, a global optimization algorithm, can be excessive. However, the application of Markov random field priors to Bayesian estimation still promises a dramatic benefit. Therefore, interest has been directed towards deterministic MAP algorithms which require fewer computations though they only achieve local optima. Most applications of Markov random field models in image processing have focused on the linear Gaussian data model where the image pixels are restricted to a small number of integer values. The computational demands of most algorithms increase significantly with the number of allowed values. A

maximum a posteriori (MAP) algorithm for use with Gibbs priors using a Poisson data model for which pixels are allowed to take any non-negative finite value is derived based upon a generalized expectation-maximization (GEM) formulation.

(3) In the use of Gibbs distributions for image processing, very little attention has been given to the available flexibility in defining the potential functions. This is mainly due to the excellent results which have been obtained using the simplest potential function in conjunction with line processes. The approach of incorporating the effect of the line process in the potential function of the Gibbs prior is investigated. The benefit of such an approach is a reduction in computation and in algorithm complexity. Three candidate potential functions for use in a general purpose Markov random field prior for images are compared.

(4) A Gibbs distribution has a very important parameter which specifies the degree to which the modes of the distribution are accentuated. This parameter is synonymous with the inverse of the regularization parameter in regularization. In practice, this parameter is often selected on the basis of trial and error. A statistic based approach to selection of the Gibbs prior parameters is presented. The resulting problem is formulated in a control systems framework and a simple algorithm for use in conjunction with deterministic MAP algorithms for simultaneous parameter selection and image restoration or reconstruction is proposed.

(5) In 3-D image reconstruction from stochastic projections, the computational requirements can inhibit the application of all but the most simplistic reconstruction approaches. In the interests of applying iterative Bayesian techniques

it becomes necessary to devise implementations to reduce the calculations necessary for matrix-vector multiplications. A method is presented based upon an intermediate polar pixel representation of the 3-D image so that a particular symmetry which occurs in emission tomography can be exploited. This approach maintains the resolution of the system and does not suffer from a spatially varying SNR which accompanies reconstructions performed in terms of unevenly sized pixels.

(6) There are many new emission imaging systems being proposed, some of which increase the efficiency of data collection. One system under development at this university at the time of this publication is based upon electronic collimation. This system incorporates the measurement of the Compton scattering angle of detected photons. As such, the point source response function of the system is spatially variant and the dimension of the reconstruction is unusually large. A total of 48^3 pixels are be estimated from 20.7×10^6 data values. A probabilistically based factorization of the system matrix is introduced which allows reduction of the computational demands of a class of iterative algorithms to be reduced by several orders of magnitude. An 'algebraic reconstruction technique' type modification to a well known ML algorithm is introduced which reduces the number of iterations required for convergence of the likelihood by a factor greater than five.

Chapter 2

Observation Models and Applications

2.1 Introduction

This chapter presents the observation models in which we are interested and several of the many areas in which these models can be applied. A significant amount of our focus has been in the area of emission tomography. Therefore, a more in-depth introduction to this area is included.

A Note on Notation: In this and all following chapters vectors are represented with emboldened lowercase characters and matrices with emboldened uppercase characters. Unless otherwise stated, subscripts are used to indicate a particular element of the vector or matrix. Vector superscripts are used to indicate the particular iteration at which the vector has been computed. The symbol $\boldsymbol{\lambda}$ is only used to represent a vector containing the lexicographically ordered pixels of an image and $\mathbf{y}$ is only used to represent a vector of data.

2.2 Observation Models

The most general stochastic observation model with which we are concerned characterizes the observation vector $\mathbf{y}$ by its density function conditioned on the undegraded image $\boldsymbol{\lambda}$. This conditional density function $f(\mathbf{y}|\boldsymbol{\lambda})$ may also depend on an additional parameter vector $\boldsymbol{\mu}$ so that we may write $f(\mathbf{y}|\boldsymbol{\lambda},\boldsymbol{\mu})$. The elements of vector $\mathbf{y}$ and $\boldsymbol{\lambda}$ may be constrained to the set of integer values [0,255], the set of non-negative integers, or allowed to take any non-negative values within a continuous range. Digitized images have usually been quantized to 256 grey levels 0-255.

Medical images are often quantized to a greater number of levels to allow closer inspection of image regions which would appear more uniform with fewer quantized levels. We are primarily concerned with two specific forms of this general stochastic observation model.

The first stochastic observation model is the linear normal model

$$\mathbf{y} = \mathbf{P}\boldsymbol{\lambda} + \mathbf{n} \tag{2.1}$$

where $\mathbf{y}$ is the mx1 data vector, $\boldsymbol{\lambda}$ is the nx1 vector containing the unknown image, $\mathbf{P}$ is the mxn blurring of defocusing matrix, and $\mathbf{n}$ is a multinormal noise vector $N(0,\mathbf{Q})$. The simplest case is that for which no blurring or defocusing is present and the elements of $\mathbf{n}$ are i.i.d. $N(0,\sigma^2)$ so that $\mathbf{P} = \mathbf{I}$, where $\mathbf{I}$ is the mxn identity matrix, and $\mathbf{y}$ is multinormal $N(\boldsymbol{\lambda},\sigma^2\mathbf{I})$. The next level of generality for the linear Gaussian model allows $\mathbf{n}$ to have non-diagonal covariance matrix $\mathbf{Q}$ so that $\mathbf{y}$ is $N(\mathbf{P}\boldsymbol{\lambda},\mathbf{Q})$. The final level of generality allows an image dependent component to the covariance matrix $\mathbf{Q}$.

The second observation model in which we are interested is the Poisson model. This observation model assumes the observations $\mathbf{y}$ to be conditionally independent Poisson variables with mean equal to a blurred version of the uncorrupted image $\boldsymbol{\lambda}$, i.e. $E\{\mathbf{y}\} = \mathbf{P}\boldsymbol{\lambda}$. In this model, the uncorrupted image $\boldsymbol{\lambda}$ is allowed to take the complete range of non-negative values. The distribution of the data $\mathbf{y}$ conditioned upon the uncorrupted image $\boldsymbol{\lambda}$ is

$$f(\mathbf{y}\,|\,\boldsymbol{\lambda}) = \prod_i e^{-(\sum_j \mathbf{P}_{ij}\boldsymbol{\lambda}_j)} \frac{(\sum_j \mathbf{P}_{ij}\boldsymbol{\lambda}_j)^{\mathbf{y}_i}}{(\mathbf{y}_i)!} . \tag{2.2}$$

The restoration and reconstruction approaches we consider are based upon these

general observation models and are applied to areas in which these models are valid. Some of these applications are discussed in the following sections.

2.3 Applications

2.3.1 Emission Tomography

Since many of the results we present in this thesis are from emission tomography systems, we will discuss this area in more detail. Emission tomography has as its major focus the localization and quantization of concentrations of radiolabeled compounds or radiopharmaceuticals introduced into a patients body. Emission tomography presents a 3-D image reconstruction problem, that of determining the 3-D distribution of either positron or gamma emitting material in the body. This is achieved through detection of gamma ray emissions from the radiolabeled compound. Emission tomography provides an important tool for diagnosis of certain diseases and for assessing the efficacy of chemotherapeutic agents. Upon introduction into a patient, a radiopharmaceutical either rapidly or gradually accumulates in various regions of the body. As the nuclei of the radioactive atoms in the compound decay, γ-rays are emitted. Some of these γ-rays exit the body and are recorded by the imaging system.

2.3.1.1 Photon Imaging

As a γ-ray propagates through the body it passes through tissues of varying densities. The probability that the γ-ray will be absorbed or Compton scattered in each of these tissues is computable as a function of the density of the tissue, the energy of the gamma ray, and the path length traveled by the γ-ray. A gamma ray

which exits the body along a path towards the gamma camera may be detected by the emission imaging system and recorded as a count.

Examining the process on the nuclear level, isotopic atoms are bonded with compounds of clinical interest and introduced into the patient. The compound localizes in various regions of the patient. The nucleus of each isotopic atom has a certain probability of decay within any time interval. The decay of a nucleus results in the emission of a gamma ray and the reduction of the energy of the nucleus to a lower state. The direction of the gamma ray is random and uniformly distributed over all possible directions in 3-D space [2].

Detection of the γ-rays is usually accomplished through the use of a gamma camera. The development of the gamma camera began in 1947, when Dr. Hofstadter constructed and tested the first NaI scintillation counter at Princeton University. In 1956, Dr. H.O. Anger developed the the gamma ray camera in the form that is widely used today. The gamma camera has three main components [2] (figure 2.1): a scintillation crystal (a transluscent material which gives off light when struck by gamma radiation), an array of photomultiplier tubes (light sensitive devices whose output is proportional to light input), and a circuit for combining the output of the photomultiplier tubes to produce an estimate of the x,y coordinates of the pulse of light generated by a gamma ray incident on the crystal. A collimator, generally a lead plate with very small holes bored through it, restricts the passage of γ-rays to the scintillation crystal unless their incident direction carries them through the drilled holes.

A γ-ray imaging system may be considered to have several hundred thousand separate bins within which it may collect and count gamma rays. The number of counts in each bin comprise the data. Corresponding to each bin is a relatively

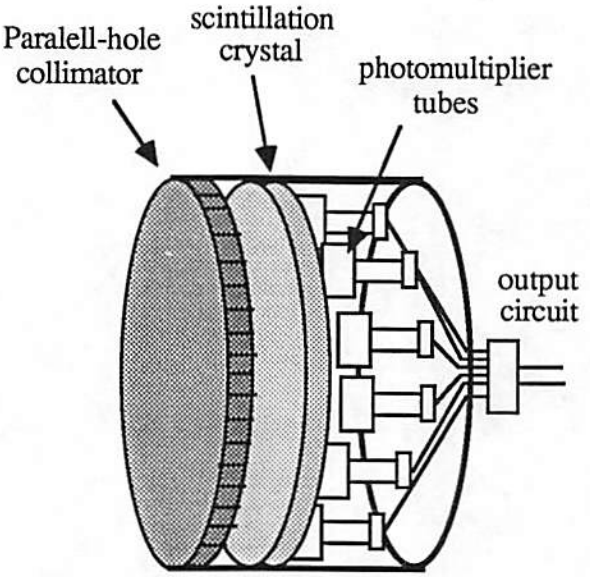

Figure 2.1
Diagram of the main components of a gamma camera.

small 3-D volume from which, with high probability, the recorded emission originated. However, these 3-D volumes overlap one other so that any point in the source space may belong to 50 or more bins. Typically, several million γ-rays will be detected. The number of counts in these bins can be shown to be conditionally independent Poisson random variables [35],[64]. To determine the distribution of radiopharmaceutical, we divide the 3-D space viewed by the imaging system into small volumes called pixels (figure 2.2). For each pixel, a quantity which characterizes the amount of isotope in that pixel is estimated. We can concatenate those pixel values into a vector $\boldsymbol{\lambda}$ for ease of notation. For display, each plane of the 3-D space will be shown as a 2-D image, the intensity of each pixel of the 2-D image will represent the concentration of isotope within its corresponding 3-D source pixel normalized to the 3-D source pixel having the highest concentration.

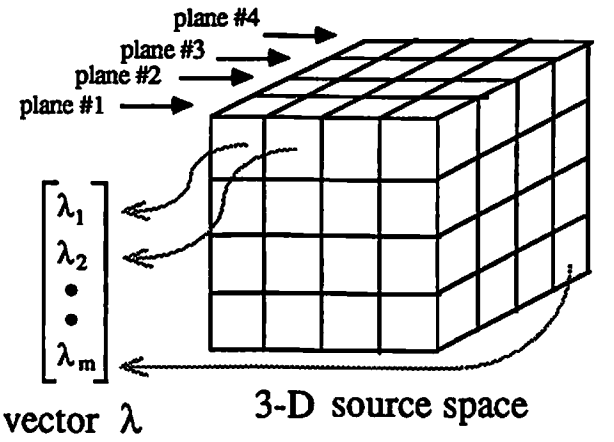

Figure 2.2
Division of a 3-D source into rectangular pixels. The concentration of isotope in each 3-D pixel volume can be concatenated into a vector λ. Display is done plane-by-plane, with each plane represented by a 2-D intensity image.

2.3.1.2 Commercial Systems

The most common emission imaging system in use today is the mechanically collimated single photon emission system [2] figure 2.3. This is a system based on a gamma camera and a lead plate which attenuates most gamma rays not incident normally to the face of the camera. Single photon emission systems are designed for use with gamma ray emitting radionuclides such as ^{99}Tc. Decay of a ^{99}Tc nucleus results in the emission of a single gamma ray in a random direction. The gamma camera registers x,y coordinates of those gamma rays which travel nearly perpendicularly towards the camera and pass through the small holes in the lead collimator. The x,y coordinates of that detection localize the emission to within a tubular volume extending perpendicularly back from the detection site. Sources of erroneous detections are gamma rays which penetrate the lead collimator without being attenuated and gamma rays which compton scatter before detection.

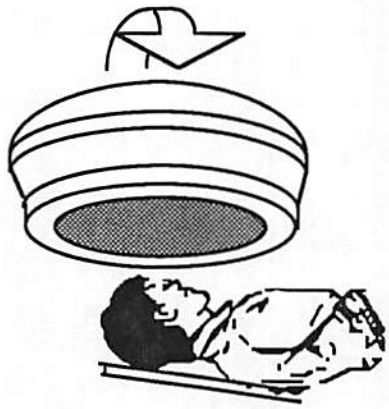

Figure 2.3
Diagram of a mechanically collimated SPECT system. Counts are caused by detection of gamma-rays passing unattenuated through a lead collimator.

One of the newer more expensive types of emission imaging systems are the positron emission tomographic (PET) systems [2]. These systems are designed for use with isotopes of carbon, oxygen, and nitrogen which decay by positron emission. The short half-life of these isotopes require a cyclotron at the PET system site, greatly increasing the cost of a complete system. However, carbon, oxygen, and nitrogen are involved in a very wide range of physiological systems and are present in many pharmaceutical compounds. A decaying nucleus of an isotopic atom emits a positron which typically travels no furthur than a few millimeters before annihilating with an electron, producing two oppositely directed gamma rays. PET imaging systems are constructed as a ring of detectors which register the simultaneous detection of gamma rays on two different detectors (figure 2.4). Noncoincident gamma ray detections are ignored. A simultaneous detection localizes the annihilation within a tubular volume between the two detectors.

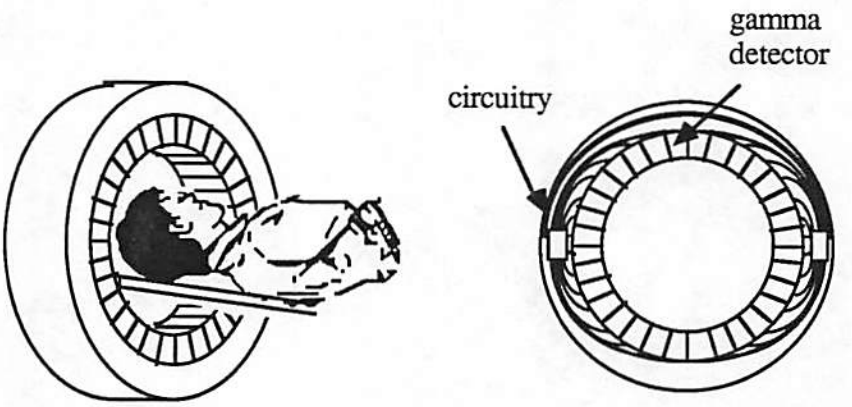

Figure 2.4
Diagram of a PET imaging system. Counts are caused by simultaneous detection of gamma-rays in two detectors.

Sources of erroneous detections are simultaneous detection of two gamma rays from two different positron/electron annihilations and compton scattering of one or both oppositely traveling gamma rays before detection. Sources of uncertainty in a detection are the distance a positron travels from the nucleus before annihilating with an electron and the departure from colinearity of the two opposite traveling gamma rays which may amount to as much as 0.5°.

Various other photon imaging system geometries and methods have been proposed and are under development.

2.3.1.3 The Poisson Data Model

The generation of data counts in an emission imaging system depends on a sequence of random events. It is neither a deterministic function of the number of source emissions nor of the amount of isotopic material. Once the source space has been discretized, the discrete Poisson model emerges. Let us examine a time interval T over which a distribution of radioactive isotopic atoms changes very little.

Let p equal the probability that a given isotopic nucleus will decay within the time interval $t_0 \rightarrow t_0+T$. At time t_0 , a given voxel (volume element) contains a fixed number of isotope atoms, say N. Then the number of γ - rays emitted from that voxel over the time interval has a binomial distribution.

$$P(\#\text{ of }\gamma\text{ - rays emitted }=n\ ) = \binom{N}{n} p^n (1-p)^{N-n} \tag{2.3}$$

Even for minute traces of the isotope, N will be very large. In addition, the half-life of most radiactive tracers is large with respect to the interval over which the detection procedure extends (the half-life of ^{99}Tc is 6 hrs.). Therefore, p is very small. As $N \rightarrow +\infty$ and $p \rightarrow 0$ in such a way that $Np \rightarrow \lambda$ this binomial distribution approaches the Poisson distribution. Since emissions from one small volume do not affect emissions from any other small volume, the numbers of γ - rays emitted from the source voxels are well modeled as independent Poisson random variables. The joint distribution function of the data for the photon imaging systems described in the previous section can be developed from this model.

Divide the scintillation crystal of the gamma camera into pixels numbered (1,2,....M) . Restrict sources to lie within some confined volume. Divide that volume into voxels (volume elements) numbered (1,2,....N). Let

$\mathbf{e}_i$ = the number of γ -rays emitted from voxel i.

$\mathbf{y}_i$ = the # of recorded counts at pixel j from

$\mathbf{D}_{i,j}$= the number of emissions from source voxel j which were recorded at data pixel i. Here, $\mathbf{D}_{i,j}$ is an intermediate random variable which is a function of $\mathbf{e}_j$.

$\mathbf{P}_{i,j}$ = the probability that a γ - ray emitted from source voxel j will generate a count in camera pixel i. $\mathbf{P}_{i,j}$ is a function of the geometry of the system, the probability of attenuation of the γ-ray, and the location of the source pixel.

The variables $\mathbf{e}_i$ are well-modeled as independent Poisson random variables with unknown parameters λ_i i=1,...,N . Therefore,

$$P(\mathbf{e}_i=n_i \mid \lambda_i) = \frac{e^{-\lambda_i}(\lambda_i)^{e_i}}{\mathbf{e}_i!} \tag{2.4}$$

$\{\mathbf{P}_{i,j}\}$ and $\{\mathbf{y}_i\}$ are known ; $\{\mathbf{D}_{i,j}\}$ and $\{\mathbf{e}_i\}$ are unknown random variables, and the $\{\lambda_i\}$ are unknown, deterministic parameters. The data $\{\mathbf{y}_i\}$ are formed by

$$y_i=\sum_{j=1}^{N}\mathbf{D}_{i,j} \tag{2.5}$$

The probabilty mass function for the intermediate variables $\mathbf{D}_{i,j}$ is

$$P(\mathbf{D}_{i,j}=n \mid \lambda_j) = \sum_{k=1}^{\infty}P(\mathbf{e}_j=k)P(O_{n,k,j,i}) \tag{2.6}$$

where $P(O_{n,k,j,i})$ is the probability of the outcome (n of k emissions from source voxel j and are recorded at data pixel i)

$$P(d_{i,j}= n \mid \lambda_j) =\sum_{k=n}^{\infty}\frac{e^{\lambda_j}(\lambda_j)^k}{k!}\binom{k}{n}(P_{i,j})^n(1-P_{i,j})^{k-n} \tag{2.7}$$

which reduces to $$P(d_{i,j}= n \mid \lambda_j) = \frac{(\mathbf{P}_{i,j}\lambda_j)^n}{n!}e^{-(\mathbf{P}_{i,j}\lambda_j)} \tag{2.8}$$

i.e. $\mathbf{D}_{i,j}$ is Poisson with parameter $(\mathbf{P}_{i,j}\lambda_j)$. Furthur, since all the $\mathbf{e}_i$'s are independent - all the $\mathbf{D}_{i,j}$'s are independent. Therefore, the data values $\mathbf{y}_i$ are independent sums of independent poisson random variables and are thus conditionally independent Poisson random variables.

$$\mathrm{E}\{\mathbf{y}_i\} =\sum_j \mathrm{E}\{\mathbf{D}_{i,j}\} =\sum_{j=1}^{N}\mathbf{P}_{i,j}\lambda_j \tag{2.9}$$

$$P_{\mathbf{y}}(y_i) = \frac{(\mathbf{P}_{i,j}\lambda_j)^{y_i}}{y_i!} e^{-(\mathbf{P}_{i,j}\lambda_j)} \tag{2.10}$$

$$P_{\mathbf{y}}(y_1, \ldots, y_M) = \prod_{j=1}^{M} \frac{(\mathbf{P}_{i,j}\lambda_j)^{y_i}}{y_i!} e^{-(\mathbf{P}_{i,j}\lambda_j)} \tag{2.11}$$

i.e. the data y are conditionally independent Poisson random variables with $E\{\mathbf{y}\} = \mathbf{P}\lambda$.

2.3.1.4 Historical Literature Review

In the late 1960's and early 1970's , manufacturers began developing gamma camera's with digital output. This gave rise to the first papers dealing with computer-based systems. Goldman et al. [25] 1970, Ekstrom and Rater [19], and Bell, Mcclain, and Ross [3] 1972 are three examples. These first efforts were restricted to processing images from a single angle of view. Operations such as smoothing, thresholding, and data correction were performed on the 2-D images representing the number of gamma rays normally incident on the camera over a set counting time.

1973 and 1974 saw a developing interest in 3-D reconstruction of emission profiles from scans over multiple angles. In 1974, a number of approaches were published. Budinger and Gullberg [10] and a number of other investigators examined filtered back-projection and linear system approaches such as the algebraic reconstruction technique ART [29]. In 1976, Rockmore and Macovski introduced a maximum likelihood approach [60]. Shepp and Vardi [64], 1982, applied the expectation - maximization approach of Dempster, Laird and Rubin [18] to formulate the popular EM algorithm for estimation of the mean emission process from positron emission data. Snyder and Miller [71] 1985 use a regularization approach

of Grenander [30]. Geman and D. McClure [23] 1985 presented their Bayesian approach to emission tomography based on a Markov random field prior in the form of a Gibbs distribution. Many important aspects of emission tomography such as attenuation compensation [12] have been addressed by various investigators.

2.3.2 Ultrasound and Geophysical Imaging

One application of the linear Gaussian observation model is in ultrasound imaging which is a non-invasive method for determining an acoustic profile of anatomical regions within the body. This method is of diagnostic importance because many types of healthy tissues and abnormal tissues can be uniquely characterized by their acoustic properties. Ultrasound imaging uses an electro-acoustic transducer to transfer electrical pulses into high-frequency acoustic waves and high-frequency acoustic waves into electrical signals. An electrical pulse generates an acoustic pulse from the tranducer which penetrates the body and is partially reflected by impedance mismatches between tissue layers. The transducer receives the reflected acoustic waves and responds with an electrical signal. In commercial systems using a single transducer, the direction in which the acoustic pulse is sent is swept back and forth so that reflections from within a fan-shaped region of a plane are received [38].

In ultrasound imaging, the form of the generated acoustic pulse is highly stable from one pulse to the next. This pulse can be measured and recorded by sending an acoustic pulse into a tank containing a homogeneous medium and receiving the reflected pulse from a plexiglass sheet at the bottom of the tank. Under certain modeling assumptions, the reflected signal from tissues layers within the body can be modeled as a superposition of reflected pulses plus additive noise. These

modeling assumptions are that the acoustic pulses are normally incident upon tissue layers and that the velocity of sound is very nearly the same in all soft body tissues. In regions which contain no bone, these modeling assumptions are well founded [39]. Let vector y contain the time-sampled signal from the transducer over the intervals following generated acoustic pulses. These observations correspond to reflections from within a fan-shaped 2-D region. The reflection coefficients from that fan-shaped 2-D region form a fan-shaped 2-D image which can be concatenated into a vector λ. The observation vector y can be related to the 2-D reflection coefficient image λ by a linear transformation **P** containing attenuated samples of the acoustic pulse, plus additive uncorrelated Gaussian noise. The 2-D maps of reflection coefficient could be expected to exhibit local dependencies and can be modeled as realizations of a Markov random field.

Reflection seismology presents a very similar problem. In this case, a map of the acoustic properties of a region below the earths surface is desired in order to draw conclusions regarding the presence or absence of oil or mineral deposits. Under similar modeling assumptions, the problem can be formulated in like manner [44]. The 2-D or 3-D acoustic maps can be expected to consist generally of layers, with impedance mismatches between the layers. The Markov random fields models used in this thesis can be used for this application.

2.3.3 Photon Counting Astronomical Imaging

In the field of astronomy, multichannel photon counting detectors are used for spectroscopic observation of faint sources [59]. Two dimensional images are formed by binning and counting photons emitted by 3-D stellar configurations millions of light-years away. These photon count images are degraded by the effects of Poisson

noise and are blurred by the detection of photons arriving from outside the field of view of each individual detector. The recorded image at the telescope can be represented in a vector $\mathbf{y}$. Each element y_i represents the number of photons arriving from a small sector of the sky. The undegraded image $\boldsymbol{\lambda}$ contains the mean emission rate from these sectors. The detection of photons from outside the field of view of each detector element can be represented by a blurring $\mathbf{P}$ of the undegraded image. Vector $\mathbf{y}$ consists of conditionally independent Posson random variables with mean $\mathbf{P}\boldsymbol{\lambda}$. The undegraded image of the map of photon sources in the night sky can be modeled as a Markov random field. The 2-D observation images can be processed by algorithms developed in this thesis.

Chapter 3

Probabilistic Solutions

3.1 Introduction

A statistic is defined as a measure computed from a set of random observations. A statistic may be descriptive, in that it characterizes a set of data, or it may be inductive, in that it is used to make an inference about the population from which the observations were drawn. In this work, we are interested in the inductive use of statistics. A statistic is a transformation or mapping from the space of all sample points to the values of a function. In this sense, the likelihood function and Bayes function are statistics as are their maxima. We therefore may characterize ML and MAP estimation as statistical estimation approaches. This viewpoint leads to the use of additional statistics for stopping criterion and parameter selection presented in chapters 4 and 6. Section 3.2 introduces the ML estimation, regularization, and Bayesian estimation approaches. Their common properties and form are pointed out in section 3.2.4. Section 3.3 introduces the spatial interaction models which are used in Bayesian estimation in later chapters. Section 3.4 discusses some of the more recent algorithms being applied to Bayesian image processing.

3.2 Maximum Likelihood, Regularization, and MAP Estimation

3.2.1 Maximum Likelihood

Suppose that the generation of data vector $\mathbf{y}$ is described by a model given by a probability function $f(\mathbf{y}|\boldsymbol{\lambda})$ where $\boldsymbol{\lambda}$ is an unknown vector of parameters. For a given data vector $\mathbf{y}$, the likelihood function $L_{\mathbf{y}}(\boldsymbol{\lambda})$ is defined as any function of $\boldsymbol{\lambda}$ which is proportional to $f(\mathbf{y}|\boldsymbol{\lambda})$.

It is a reasonable and accepted notion that, for a given $\mathbf{y}$, the likelihood $L_{\mathbf{y}}(\boldsymbol{\lambda})$ is a measure of the degree to which the data support any value of $\boldsymbol{\lambda}$. In fact, given an observed $\mathbf{y}$, the likelihood function contains all *experimental* information which is relevant to inferences about $\boldsymbol{\lambda}$. A natural application of this concept is the estimation approach called maximum likelihood [50],[72] which accepts as the estimate of $\boldsymbol{\lambda}$ that $\hat{\boldsymbol{\lambda}}$ which maximizes the likelihood function. If there is no unique maximum, then any $\hat{\boldsymbol{\lambda}}$ which achieves the maximum is called a maximum likelihood state.

We can think of the likelihood function, $L_{\mathbf{y}}(\boldsymbol{\lambda})$ as taking on different values for different realizations of the random data vector $\mathbf{y}$. That is, $L_{\mathbf{y}}(\boldsymbol{\lambda})$ assigns an entire function to every outcome $\mathbf{y}$. In this sense, the likelihood function is a statistic. Likewise, an outcome $\mathbf{y}$ defines the maximum likelihood states so that ML estimates are themselves statistics.

Often, an investigator has some additional insights into $\boldsymbol{\lambda}$ besides that derived solely from the data. We would like some basis for combining these insights with the data to arrive at an estimate. A more urgent situation occurs in an application for which the ML estimate does not possess some quality or structure essential to a reasonable solution of the problem. In image processing, it is held that images are

well represented as consisting of segmented, relatively smooth regions. Maximum likelihood estimates of noisy images do not possess this structure. Two methods of incorporating this structural insight in image reconstruction are regularization and Bayesian estimation.

3.2.2 Regularization

The problem of restoring or reconstructing an image from a set of image dependent data can be formulated by posing a model for the data y and a measure $h(\mathbf{y},\hat{\boldsymbol{\lambda}})$ for evaluating an estimate $\hat{\boldsymbol{\lambda}}$ based upon the data. Such a measure $h(\cdot,\cdot)$ may be a likelihood function or some function of the error between the data and the output of the data model based an estimate $\hat{\boldsymbol{\lambda}}$. The reconstruction is thus formulated as an optimization problem

$$\max_{\boldsymbol{\lambda}} \; h(\mathbf{y},\boldsymbol{\lambda}) \tag{3.1}$$

Two problems which often arise when a reconstruction is based on a single measure $h(\mathbf{y},\boldsymbol{\lambda})$, such as the likelihood, are that the solution may not be unique or that the solution may not possess some expected quality such as smoothness. One approach to these two problems is to impose an additional measure $g(\hat{\boldsymbol{\lambda}})$ which is solely a function of the image estimate $\hat{\boldsymbol{\lambda}}$ and to solve the optimization problem

$$\max_{\boldsymbol{\lambda}} \; h(\mathbf{y},\boldsymbol{\lambda}) + \gamma\, g(\boldsymbol{\lambda}) \qquad 0 < \gamma < +\infty \tag{3.2}$$

Such approaches are referred to as regularization procedures [75],[77]. As $\gamma \rightarrow 0$ the solution becomes entirely dependent on the first measure $h(\mathbf{y},\boldsymbol{\lambda})$ and as $\gamma \rightarrow +\infty$ the solution becomes independent of the data y. Values of γ between zero and infinity represent trade-offs between the solutions corresponding to those regularization

parameter values.

For example, consider the linear data model $\mathbf{y} = \mathbf{P}\boldsymbol{\lambda} + \mathbf{n}$ where $\mathbf{n}$ is an noise vector with $E\{\mathbf{n}\} = 0$ and $E\{\mathbf{n}\mathbf{n}^t\} = \sigma^2\mathbf{I}$. If the noise vector is additionally Gaussian the least squares solution $\boldsymbol{\lambda} = (\mathbf{P}^t\mathbf{P})^{-1}\mathbf{P}^t\mathbf{y}$ is also the maximum likelihood estimator. This estimate may be unstable due to the influence of the noise vector $\mathbf{n}$ and poor condition number of matrix $\mathbf{P}^t\mathbf{P}$. A regularized approach would be to reformulate the problem as

$$\min_{\boldsymbol{\lambda}} \| \mathbf{y}-\mathbf{P}\boldsymbol{\lambda} \|_2^2 + \gamma \| \boldsymbol{\lambda} \|_2^2 \tag{3.3}$$

The unique regularized solution is then

$$\hat{\boldsymbol{\lambda}} = [\mathbf{P}^t\mathbf{P} + \gamma\mathbf{I}]^{-1}\mathbf{P}^t\mathbf{y} \, . \tag{3.4}$$

In the statistical literature, such an approach is termed ridge regression [37]. In this terminology $\mathbf{y} = \mathbf{P}\boldsymbol{\lambda} + \mathbf{n}$ is the regression model where vector $\mathbf{n}$ is simply the regression error. A ridge estimator in its simplest form is expressed as in equation (3.4) where γ is some non-negative scalar constant. This estimator always exists for any $\gamma > 0$ and has lower mean squared error for values of γ within a range as specified by Hoerl and Kennard [37]. A large number of publications have addressed this class of estimators.

For regularization, a variety of choices for $g(\boldsymbol{\lambda})$ have been examined in image processing applications [77]. Many can be represented as

$$g(\boldsymbol{\lambda}) = \boldsymbol{\lambda}^t\mathbf{K}\boldsymbol{\lambda} \tag{3.5}$$

where $\mathbf{K}$ is some appropriate positive definite matrix chosen to emphasize the local smoothness of an image. One example is

$$g(\lambda)=\sum_{i}\sum_{j\varepsilon N_i}(\lambda_i-\lambda_j)^2 \tag{3.6}$$

where N_i is the set of indices of pixels in the region local to λ_i. The measure $g(\lambda)$ can also be formulated to reflect discrete derivatives or measures on the normalized image [77].

Selection of the regularization parameter γ is often done through trial and error or else based upon some knowledge of the SNR of the image. This is discussed further in chapter 6.

3.2.3 Bayesian Estimation

Bayesian estimation [4] is another method of incorporating into the estimate some insight into the structure of the unknown parameter λ. Bayesian estimation requires specification of a suitable prior probability function for λ. The probability distribution of the image vector λ conditioned on the data vector y can then be formulated using Bayes rule

$$f(\lambda \mid y) = \frac{f(y \mid \lambda) f(\lambda)}{f(y)} \tag{3.7}$$

where $f(y)$ is a constant for a given data. The posterior distribution $f(\lambda \mid y)$ contains all information, both experimental and prior, about λ.

Three common Bayesian estimates of λ are the mode, the mean, and the median of the posterior distribution, equation (3.7). The maximum a posteriori (MAP) estimate is sometimes refered to as generalized maximum likelihood estimate since it is the estimate $\hat{\lambda}$ which maximizes the posterior distribution $f(\lambda \mid y)$ rather than the likelihood. Since $f(y)$ is a constant for a given data vector, MAP reconstruction requires solution of the problem

$$\max_{\lambda} B(\lambda \mid y) = \log f(y \mid \lambda) + \log f(\lambda) \tag{3.8}$$

The posterior mean and median are also common Bayesian estimates. If the posterior is symmetric and unimodal; then the mode, mean, and median coincide. Occasionally, the mean and median are better estimates than the mode, especially in regard to robustness with repect to changes in the prior. Bayesian approaches can also be formulated by specifying an error or loss function and selecting the $\hat{\lambda}$ which minimizes the expected value of that function with respect to the posterior distribution.

3.2.4 Similarities

Maximum likelihood estimation can be defined as a special case of MAP estimation. Consider the case where the prior distribution $f(\lambda)$ is chosen as uniform over the acceptable parameter set. Then $f(\lambda)$ is simply a constant and from equation (3.8) MAP estimation and ML estimation are equivalent.

Although regularization is a deterministic approach, it is interesting to note that a Bayesian interpetation of certain forms of regularization exists [77]. Consider a regularization parameter γ and a regularizing function $g(\lambda)$ which takes on finite values for all λ and for which

$$\int_{\lambda \varepsilon \Omega} \exp[\gamma g(\lambda)]\, d\lambda < +\infty \,. \tag{3.9}$$

If we choose, in equation (3.1), $h(y,\lambda) = \log f(y \mid \lambda)$, the conditional density function, then regularization, posed as

$$\max_{\lambda} \log f(y \mid \lambda) + \gamma g(\lambda) \tag{3.10}$$

is equivalent to

$$\max_{\lambda} f(y|\lambda) \frac{1}{C_2} e^{\gamma g(\lambda)} \tag{3.11}$$

which is MAP estimation with prior $\frac{1}{C_2} e^{\gamma g(\lambda)}$ where C_2 is the normalizing constant for the distribution. In the context of Bayesian estimation, the regularization parameter γ is equivalent to $\frac{1}{\beta}$, the inverse of the Gibbs prior parameter.

This interpetation suggests that a Bayesian approach may yield a formulation which:

(1) more closely explains the mechanism of many problems, that of a random instead of a deterministic configuration of the unknown image.

(2) provides some insight into the meaning and choice of parameters.

(3) offers a richer environment within which to formulate the insights of the investigator.

MAP estimation offers a potential for reconstruction improvement when prior distributions, more meaningful than the uniform distribution, can be defined. In image restoration and reconstruction, priors are sought which lend increased probability to realizations which feature segmented, slowly changing regions and decreased probability to highly erratic images. Such image models should incorporate local interactions while allowing abrupt changes across edges or region boundaries. Markov random field models, which we examine in the following section, have been demonstrated to be useful in this application.

3.3 Markov Random Field Models

A lattice is a collection of locations on a 2-D plane. Each location within a lattice is refered to as a site. In image processing we are interested in regular lattices in which sites correspond to the pixels of an image. A discrete Markov random field (MRF) defined on a lattice is a collection of random variables, corresponding to the sites of the lattice, for which the probability of a given site value conditioned on the values of all other sites in the lattice is equal to the probability of the site value conditioned on the values at a small subset of the lattice sites. This subset of the lattice sites which surrounds the site is called the neighborhood of the given site. Let us label all the sites in a lattice with a set of consecutive indices and let the set of indices of sites in the neighborhood of pixel j be denoted N_j. Then a Markov random field [83] is one for which

$$P(\lambda_j | \lambda_i : i \neq j) = P(\lambda_j | \lambda_i : i \in N_j) \; . \tag{3.12}$$

If λ_i is a neighbor of λ_j , then λ_j is required to be a neighbor of λ_i.

A neighborhood structure is said to be causal if there exists a lexicographic labeling for which the labels of the neighborhood sites of each pixel are less than the integer label of that pixel. If such a labeling does not exist, the neighborhood structure is called non-causal. Non-causal neighborhoods are referred to as 0^{th} order, 1^{st} order,.., N^{th} order. Figure 3.1 shows the 0^{th} order, 1^{st} order, and 2^{nd} order neighborhoods for the 2-D lattice and the 1^{st} order neighborhood for the 3-D lattice.

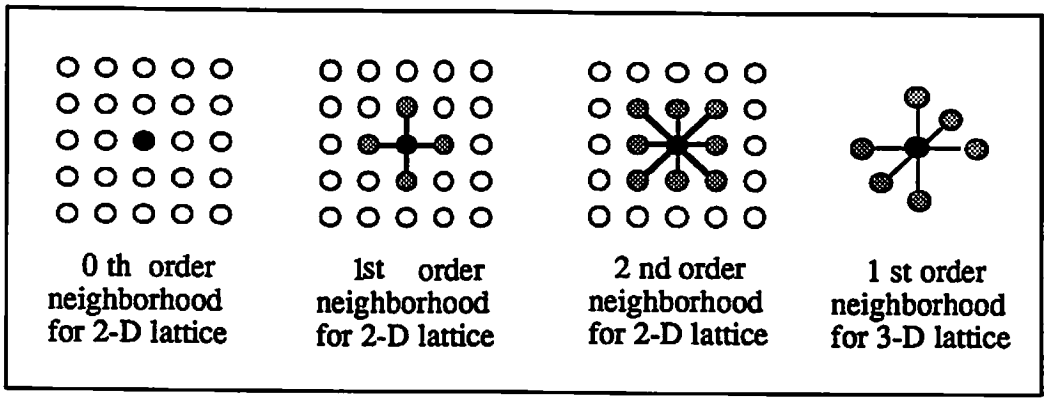

Figure 3.1
0^{th} order, 1^{st} order, and 2^{nd} order neighborhoods for a 2-D lattice and the 1^{st} order neighborhood for a 3-D lattice.

3.3.1 Gauss Markov Random Field Model

A Markov random field for which the variables have a jointly normal distribution is called a Gauss Markov random field. Let s be a two dimensional vector containing an (x,y) location on a 2-D grid and $\lambda(s)$ denote the value of the observation at s. Define a neighborhood structure letting N_s denote a symmetric set of indices of neighboring pixels s; i.e. if $s+r \in N_s$ then $s-r \in N_s$. Assume the observations obey the difference equation

$$\lambda(s) = \sum_{r \in N_s} \theta_r \left[\lambda(s+r) + \lambda(s-r)\right] + n(s) \; . \tag{3.13}$$

where θ_r is a deterministic parameter vector and $n(s)$ is a stationary Gaussian noise sequence for which

$$\begin{aligned} E\{n(s)n(s+r)\} &= -\theta_r\sigma^2 \qquad (s+r) \in N_s \\ &= \sigma^2 \qquad r=0 \\ &= 0 \qquad \text{otherwise.} \end{aligned} \tag{3.14}$$

Then it can be shown that λ is a Gauss Markov random field [83]. The visual characteristics of a sample realization of such a field is determined by the parameter vector θ_r, σ^2, and the symmetric neighbor set N_s.

3.3.2 Cliques and Gibbs Priors

Another powerful representation for a Markov random field is by a Gibbs distribution. By the Hammersly-Clifford theorem (1971) [6] a random field defined on a lattice is a Markov random field if and only if its distribution function corresponds to a Gibbs function. To define the form of a Gibbs function, we must first define a clique. For a specified neighborhood structure (0^{th} order, 1^{st} order, etc.), a clique is either a single site or a set of sites such that each site in the clique is a neighbor of all other sites in the set. The clique types associated with each neighborhood are shown in figure 3.2. For a rectangular 2-D lattice and the 1^{st} order neighborhood, the cliques are sets of sites consisting of a single site or two horizontally or vertically adjacent sites.

A Gibbs distribution is a probability measure on the set of configurations $\{\lambda\}$ which has the form

$$f(\lambda) = \frac{1}{K_\beta} e^{\frac{-U(\lambda)}{\beta}} \tag{3.15}$$

where β is the Gibbs prior parameter, K_β is the normalizing constant (partition function), and $U(\lambda)$ is termed the energy function. The energy function has the form

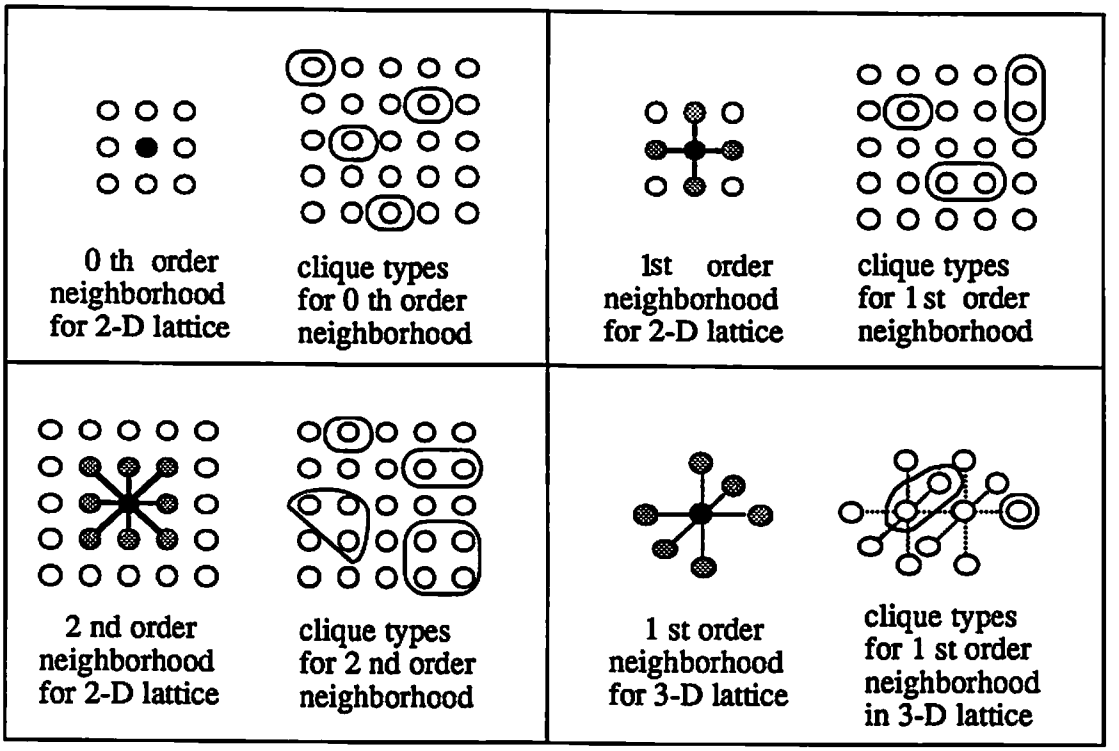

Figure 3.2
1st and 2nd order neighborhood clique types.

$$U(\lambda) = \sum_{c \varepsilon C} \sum_{I_c} V_c(\lambda_i \ldots \lambda_m : \{i..m\} \varepsilon I_c) \tag{3.16}$$

where C denotes the set of all cliques types, I_c a set of indices for the lattice sites contained in a single clique of type c, and $V_c(\lambda)$, termed a potential function, is a function of the values at the sites in I_c.

To avoid confusion, it is worthwhile to examine the Hammersly-Clifford theorem in detail as described in [6]. Let $P(\lambda)$ simply denote the probability of configuration λ occurring where $\lambda \varepsilon \Omega$ for some feasible domain Ω. Without loss of generality we will assume $0 \varepsilon \Omega$ so that $P(0) > 0$. Then,

$$Q(\lambda) = \ln\{P(\lambda)/P(0)\} \tag{3.17}$$

is well defined for any $\lambda \in \Omega$. For example, $Q(0) = 0$. Also, where $\lambda = [\lambda_1 \lambda_2 .... \lambda_N]^t$, let λ^{-i} denote the configuration $\lambda^{-i} = [\lambda_1 .... \lambda_{i-1}\, 0\, \lambda_{i+1} .... \lambda_N]^t$; i.e. λ^{-i} is the same as configuration λ with the i^{th} pixel value of λ replaced with a 0 value. Now

$$\exp\{Q(\lambda) - Q(\lambda^{-i})\} = \frac{P(\lambda)}{P(\lambda^{-i})} = \frac{P(\lambda_i \mid \lambda_1 .... \lambda_{i-1} \lambda_{i+1} .... \lambda_N)}{P(0 \mid \lambda_1 .... \lambda_{i-1} \lambda_{i+1} .... \lambda_N)}. \tag{3.18}$$

In order to satisfy the Markov property in equation (3.12), $Q(\lambda) - Q(\lambda^{-i})$ can only depend upon λ_i and its neighbors. The Hammersly-Clifford theorem in effect uses the realization that there exists an expansion of $Q(\lambda)$, unique on Ω of the form

$$Q(\lambda) = \sum_{1 \le i \le N} \lambda_i G_i(\lambda_i) + \sum_{2 \le j \le N} \sum_{i<j} \lambda_i \lambda_j G_{ij}(\lambda_i \lambda_j) + \tag{3.19}$$

$$+ \sum_{3 \le k \le N} \sum_{j<k} \sum_{i<j} \lambda_i \lambda_j \lambda_k G_{ijk}(\lambda_i \lambda_j \lambda_k) + \ldots + \lambda_1 \cdots \lambda_N G_{1..N}(\lambda_1 .. \lambda_N).$$

Consider $\lambda^{-1} = [0, \lambda_2, .., \lambda_N]^t$. From equation (3.19)

$$Q(\lambda) - Q(\lambda^{-1}) = \lambda_1 \Big\{ G_1(\lambda_1) + \sum_{2 \le j \le N} \lambda_j G_{1j}(\lambda_1 \lambda_j) + \tag{3.20}$$

$$+ \sum_{3 \le k \le N} \sum_{j<k} \lambda_j \lambda_k G_{1jk}(\lambda_1 \lambda_j \lambda_k) + \ldots + \lambda_2 \cdots \lambda_N G_{1..N}(\lambda_2 .... \lambda_N) \Big\}$$

Suppose pixel λ_q is not a neighbor of pixel λ_1. Then $Q(\lambda) - Q(\lambda^{-1})$ must not be a function of λ_q for all $\lambda \in \Omega$. Setting $\lambda_i = 0$ for all $i \neq 1$ or q, we see that $G_2(\lambda_1 \lambda_q) = 0$ on $\lambda \in \Omega$. By similar choices for λ, it can be shown that all $G_\cdot(\cdot)$ functions involving both λ_1 and λ_q must equal zero. A similar result holds for all other pixel sites which are not neighbors of site 1.

Alternatively, any set of functions $G_\cdot(\cdot)$ defines a valid probability distribution. Since $Q(\lambda) - Q(\lambda^{-i})$ depends upon λ_j only if there is a non-zero'd $G_\cdot(\cdot)$ involving

both λ_i and λ_j, from equation (3.20) the same holds for $P(\lambda_i | \lambda_1 .... \lambda_{i-1} \lambda_{i+1} .... \lambda_N)$. By equating terms, for example $\frac{1}{\beta} V_{ijk}(\lambda_i \lambda_j \lambda_k) = \lambda_i \lambda_j \lambda_k G_{ijk}(\lambda_i \lambda_j \lambda_k)$, the the derivation is completed. In choosing the $G_{\cdot}(\cdot)$ functions, it is only necessary to ensure that

$$\sum_{\lambda \varepsilon \Omega} \exp[Q(\lambda)] \tag{3.21}$$

is finite if λ can take only discrete values, or that $\exp[Q(\lambda)]$ is integrable over Ω if λ can take a continuous range of values.

For example, if we select a 1st order neighborhood, and we set the potential function defined on cliques containing a single pixel to zero, the energy function only contains a potential function for cliques containing two pixels and has the form

$$U(\lambda) = \sum_{I_{c_2}} V_{c_2}(\lambda_i, \lambda_j : \{ij\} \varepsilon I_{c_2}) \tag{3.22}$$

where c_2 is the clique type containing only two lattice sites and I_{c_2} contains a pair of indices for two lattice sites in a clique.

Gibbs functions provide a powerful class of locally dependent priors. An appeal of the Gibbs prior is that it can be specified to within a normalizing constant by defining a suitable pixel neighborhood and potential functions on the cliques associated with that neighborhood. The intent is to capture the desired properties of the unknown image by a suitable choice of energy function.

Pixel configurations of lowest energy are of highest probability density. In image processing applications, it is common to choose energy functions which penalize configurations with neighboring pixels differing by large amounts. The Gaussian prior with a diagonal covariance matrix **H** and mean image **m** is also a

particular case of a Gibbs prior with a 0^{th}-order neighborhood. A 0^{th}-order neighborhood has only cliques containing a single pixel. The corresponding energy function is $-\frac{1}{2}(\lambda-\mathbf{m})'\mathbf{H}(\lambda-\mathbf{m})$.

3.3.3 Gibbs Priors and MAP Estimation

For Bayesian image restoration or reconstruction, the normalizing constant of the Gibbs prior K_β, which depends on β, need not be calculated. The parameter β controls the degree to which the modes of the Gibbs prior are accentuated. As $\beta \rightarrow +\infty$ the Gibbs distribution tends to the uniform distribution. As $\beta \rightarrow 0$ the prior becomes increasingly more pronounced about its modes. With a Gibbs prior, (1) becomes

$$\max_{\lambda} \; B(\lambda \mid y) = \log f(y \mid \lambda) - \frac{1}{\beta} \sum_{c \in C} \sum_{I_c} V_c(\lambda_i .. \lambda_m : \{i..m\} \in I_c) \qquad (3.23)$$

so that as $\beta \rightarrow +\infty$ Bayesian reconstruction is unaffected by the prior and reduces to maximum likelihood reconstruction. In [5] and [22], acceptable values for β were obtained by trial and error. In chapter 6, it is shown that β values from the interval $+\infty$ to some lower limit value can produce an improvement in image reconstructions. Smaller β values produce a degradation due to an over-influence by the prior. Chapter 6 contains a statistical approach to selection of the prior parameter.

3.4 MAP Algorithms

In this section we discuss two deterministic MAP algorithms and a stochastic algorithm for image recontruction based on the specification of a prior distribution. The stochastic algorithm has the ability to climb out of local extremum. Simulated annealing (section 3.4.4) will in fact escape all local extremum and converge

towards the globally optimal solution given a sufficient number of iterations. The deterministic algorithms have the desireable property of faster convergence, but they may get stuck at a local extrema. There are various applications where computation time may not be much of a consideration, while other applications may place a high degree of importance upon achieving fast results.

3.4.1 Deterministic Algorithms

3.4.1.1 Iterated Conditional Modes

The iterated conditional modes (ICM) algorithm [5] is a deterministic algorithm for MAP estimation. This approach updates a single pixel at a time by maximizing a univariate function which is conditioned on all data values to which that pixel contributes, all pixels which contribute to those data values, and all neighbors of the pixel to be updated. Let $\boldsymbol{\lambda}$ be the vector containing the uncorrupted image and $\mathbf{y}$ the vector containing the blurred and noisy image. The ICM algorithm updates a single pixel of that image at a time, while holding all other pixels fixed. Let the pixel being updated be λ_i. Then, λ_i is updated to maximize $f(\lambda_i \mid \mathbf{y};\lambda_j : j \neq i)$. Since

$$f(\boldsymbol{\lambda} \mid \mathbf{y}) = f(\lambda_i \mid \mathbf{y};\lambda_j : j \neq i)\, f(\lambda_j : j \neq i \mid \mathbf{y}) \tag{3.24}$$

and $f(\lambda_j : j \neq i \mid \mathbf{y})$ is not a function of λ_i, by maximizing $f(\lambda_i \mid \mathbf{y};\lambda_j : j \neq i)$ with respect to λ_i one also maximizes $f(\boldsymbol{\lambda} \mid \mathbf{y})$ with respect to λ_i . The ICM algorithm is therefore equivalent to coordinate descent [52] of the negative conditional density function. As such, this algorithm will converge to some local minimum.

For image restoration in which the image has been degraded by a blurring function or in image reconstruction applications, each undegraded image pixel λ_i may contribute to many observation values y. In this case, maximization of $f(\lambda_i \mid \mathbf{y};\lambda_j : j \neq i)$ with respect to λ_i may be quite computationally demanding because the number of elements y_k and $\lambda_j : j \neq i$ contributing to $f(\lambda_i \mid \mathbf{y};\lambda_j : j \neq i)$ may be very large. Where blurring is very limited or non-existant, ICM, equivalently coordinate ascent, tends to converge after a relatively few iterations and is simple to implement. In such a situation, ICM may be an effective optimization approach.

3.4.1.2 Generalized EM Approach to MAP Restoration and Reconstruction

The EM algorithm as presented by Dempster, Laird, and Rubin [18] is a general approach to iterative optimization of likelihood or posterior distribution functions when the data can be formulated in a complete/incomplete framework. A complete/incomplete data formulation is applicable when data is missing or when the problem has a more natural formulation in terms of a set of unobserved data. At each iteration, the EM approach requires two steps: an expectation step (E-step) followed by a maximization step (M-step). Often, these two steps can be combined into one.

Let x be the vector of complete, but unobserved, data and y the vector of incomplete, but observed, data. In order to apply the EM approach, the relationship between the complete data and the incomplete data must be a many-to-one mapping from $\{\mathbf{x}\}$ to $\{\mathbf{y}\}$. That is, given a realization $\hat{\mathbf{x}}$, only one particular realization $\hat{\mathbf{y}}$ has non-zero probability of having occurred. Given a realization $\hat{\mathbf{y}}$, there is a feasible set $\{\mathbf{x}\}_{\hat{\mathbf{y}}}$ which has non-zero probability of having occurred.

It then follows, that due to this many x to one y mapping

$$f(\mathbf{x}\,|\,\mathbf{y}\lambda) = \frac{f(\mathbf{x}\mathbf{y}\,|\,\lambda)}{f(\mathbf{y}\,|\,\lambda)} = \frac{f(\mathbf{x}\,|\,\lambda)\, I_{\mathbf{y}}(\mathbf{x})}{f(\mathbf{y}\,|\,\lambda)} \tag{3.25}$$

where $I_{\mathbf{y}}(\mathbf{x})$ is the indicator function which is equal to 1 if x results in y and equal to 0 otherwise. In addition, for any λ^k

$$E_{\mathbf{x}}\{\log f(\mathbf{y}\,|\,\lambda)\,|\,\mathbf{y}\lambda^k\} = \int_{\{\mathbf{x}\}_{\mathbf{y}}} \log f(\mathbf{y}\,|\,\lambda) f(\mathbf{x}\,|\,\mathbf{y}\lambda^k) d\mathbf{x} = \log f(\mathbf{y}\,|\,\lambda) \tag{3.26}$$

Combining equations (3.25) and (3.26) gives

$$\log f(\mathbf{y}\,|\,\lambda) = E_{\mathbf{x}}\{\log f(\mathbf{x}\,|\,\lambda)\,|\,\mathbf{y}\lambda^k\} - E_{\mathbf{x}}\{\log f(\mathbf{x}\,|\,\mathbf{y}\lambda)\,|\,\mathbf{y}\lambda^k\} \tag{3.27}$$

which is a function of λ and y. It follows that with a complete/incomplete data formulation, $E_{\mathbf{x}}\{\log f(\mathbf{x}\,|\,\lambda)\,|\,\mathbf{y}\lambda^k\}$ and $E_{\mathbf{x}}\{\log f(\mathbf{x}\,|\,\mathbf{y}\lambda)\,|\,\mathbf{y}\lambda^k\}$ are each functions of λ^k but their difference is not. Substitution of equation (3.27) into equation (3.8) results in an expression for $B(\lambda\,|\,\mathbf{y})$ given a data vector y

$$B(\lambda\,|\,\mathbf{y}) = Q(\lambda\,|\,\lambda^k) - E_{\mathbf{x}}\{\log f(\mathbf{x}\,|\,\mathbf{y}\lambda)\,|\,\mathbf{y}\lambda^k\} \tag{3.28}$$

$$\text{where } Q(\lambda\,|\,\lambda^k) = E_{\mathbf{x}}\{\log f(\mathbf{x}\,|\,\lambda)\,|\,\mathbf{y}\lambda^k\} + \log f(\lambda) \tag{3.29}$$

To clarify how the EM approach works we first note that, from Jensen's inequality [4], for any $\lambda^{k+1} \neq \lambda^k$

$$E_{\mathbf{x}}\{\log f(\mathbf{x}\,|\,\mathbf{y}\lambda^{k+1})\,|\,\mathbf{y}\lambda^k\} \leq E_{\mathbf{x}}\{\log f(\mathbf{x}\,|\,\mathbf{y}\lambda^k)\,|\,\mathbf{y}\lambda^k\}\,, \tag{3.30}$$

with equality if and only if $\log f(\mathbf{x}\,|\,\mathbf{y}\lambda^{k+1}) = \log f(\mathbf{x}\,|\,\mathbf{y}\lambda^k)$ almost everywhere [83]. It follows that a sufficient condition for $B(\lambda^{k+1}\,|\,\mathbf{y}) > B(\lambda^k\,|\,\mathbf{y})$ is $Q(\lambda^{k+1}\,|\,\lambda^k) > Q(\lambda^k\,|\,\lambda^k)$ since the second term on the right-hand side of equation (3.25) is guaranteed, from Jensen's inequality, not to decrease.

Beginning with some initial estimate $\lambda^0 > 0$ the EM algorithm for MAP estimation thus consists of the two steps:

The E–step: form $E_x\{\log f(x|\lambda)|y\lambda^k\}$ (3.31)

The M–step: solve $\max_{\lambda} Q(\lambda|\lambda^k) = E_x\{\log f(x|\lambda)|y\lambda^k\} + \log f(\lambda)$ (3.32)

If the M-step is carried out to a global maximum of the E-step, the approach is termed an EM algorithm. If the M-step is only carried out to ensure $Q(\lambda^{k+1}|\lambda^k) \geq Q(\lambda^k|\lambda^k)$ the approach is termed a generalized EM (GEM) algorithm.

We note here that a maximum of the posterior distribution, equation (3.23), is not obtained by a single M-step since the M-step only involves maximization with respect to a portion of the posterior distribution with the guarantee that the remaining portion is increased but not maximized. This is also the key to why the EM/GEM approach is not guaranteed to achieve a global maximum even if the M-step involves a global maximization [83]. As shown above and in [18], an EM/GEM approach ensures an increase in $B(\lambda|y)$ so that, for $B(\lambda|y)$ bounded from above, convergence to some B^* is assured. Continuity of $Q(\lambda|\lambda^k)$ with respect to both λ and λ^k is sufficient to ensure that all limit points of the sequence $\{\lambda^k\}$ are stationary points of $B(\lambda|y)$ [83]. In general, if $B(\lambda|y)$ is not unimodal and the set of stationary points contains points which are not local maxima, the EM/GEM approach at best only assures convergence of the sequence $\{\lambda^k\}$ to a stationary value. As Wu states, this should not be suprising since in such a case no general optimization algorithms are guaranteed to converge to local maxima. Except for a single exception, an EM algorithm has no additional convergence properties compared to a GEM algorithm. This exception, which is provided by Wu

[83], occurs when it can be demonstrated that any stationary point which is not a local maximum is additionally not a global maximum of the E-step. Since the M-step of an EM algorithm requires global maximization of the results from the E-step, the EM M-step would not arrive at a stationary point which was not a local maxima under the above condition. The difficulty in verifying this condition, were it in fact to hold true for a given problem, makes this condition mainly of theoretical interest. From a practical viewpoint, the question of whether to carry out a global maximization versus an increase within the M-step is solely one concerned with increasing the per-iteration convergence speed at an increased per-iteration computational cost. For a complete treatment of the EM/GEM approach and its convergence properties see [18] and [83].

Optimization of likelihood functions or Bayesian functions with independent priors may result in closed forms for the M-step. If the complete data $\mathbf{x}$ are independent, the complete data are a linear function of the incomplete data, and the image pixels λ are treated as independent, the M-step only requires optimization of a set of univariate functions. However, it is generally accepted that the structure of images is one of local correlation [6],[61]. For MAP estimation with an EM approach, correlated priors prohibit the existance of closed form solutions for an EM M-step. Each EM M-step thus requires an iterative optimization of an N-dimensional function, N being the dimension of λ. What results is an iterative optimization algorithm within each iteration of an iterative optimization algorithm. In this case, iterative maximization of the posterior distribution by a method such as ICM [5] without using an EM formulation would seem more sensible than an EM approach. This problem can be overcome. In chapter 5, we present a GEM algorithm which offers an attractive alternative to both of these approaches and exhibits fast convergence to a local extremum.

3.4.2 Stochastic Relaxation

Any iterative probabilistic optimization approach which permits changes to configurations of lower probability according to a predefined rule can be termed stochastic relaxation. The intent is to define the algorithm such that it generates an ensemble of configurations whose distribution converges to a desired distribution. Stochastic relaxation algorithms are typically 'site replacement' algorithms in that a single element of a 2-D or 3-D image is updated at each step. The advantage of stochastic algorithms are that they are able to escape local extrema and under certain asymptotic conditions converge towards the globally optimal solution. The Metropolis algorithm and the Gibbs sampler can be used to generate sample configurations from Gibbs distributions [16],[22] which are discussed in section 3.3.2. In addition, these algorithms are at the heart of simulated annealing, a new algorithm for combinatorial optimization. It has been shown [55] that simulated annealing gives asymptotically the optimum solution with probability 1. For these reasons it is worthwhile to examine the Metropolis algorithm in some detail.

3.4.2.1 Metropolis Algorithm

The Metropolis algorithm [53] was introduced in 1953 as a method of simulating the evolution of a solid to thermal equilibrium. Let vector λ contain the locations of the molecules of a solid in R^3, the location of each molecule corresponding to a single element of λ. Each configuration λ has associated with it an energy $E(\lambda)$ which is a function of the distances between molecules. Given a fixed temperature T, the probability of a configuration with energy $E(\lambda)$ at equilibrium is given by the Boltzmann distribution

$$f(E) = \frac{1}{Z_T}\exp(\frac{-E}{k_B T}) \tag{3.33}$$

where Z_T is a normalization factor dependent upon the temperature T and k_B is the Boltzmann constant.

The Metropolis algorithm is a 'site replacement' algorithm, in that only a single element of λ is changed at each step. Elements are visited in random order. When visiting an element λ_i, a perturbed configuration $\lambda+\Delta\lambda$ is generated by displacing λ_i by a small random amount $\Delta\lambda_i$. If the difference in energy

$$\Delta E = E(\lambda+\Delta\lambda) - E(\lambda) \tag{3.34}$$

between the perturbed configuration and the present configuration is negative, the perturbed configuration is of lower energy and more probable as an equilibrium state of the solid. If ΔE is positive, then the perturbed configuration is less probable at equilibrium. Given a perturbed configuration and the ΔE, the Metropolis algorithm accepts the perturbed configuration with probability

$$\text{Prob(acceptance)} = \min[1,\exp(\frac{-\Delta E}{T})] \tag{3.35}$$

That is, if the perturbed configuration is of lower energy, it is accepted with probability 1; if it is of higher energy, ΔE is positive and the perturbed configuration is accepted with probability $\exp(\frac{-\Delta E}{T})$. Visiting elements of λ at random, the Metropolis algorithm perturbs and updates λ according to equation (3.35).

Implementation of the Metropolis is based upon a set of uniform (0,1) pseudo-random numbers u_j. Let $\lambda \varepsilon R^M$ and let all elements of λ be restricted to some finite interval $[\gamma_1,\gamma_2]$. To select the next element of λ to be updated, generate u_i and visit element λ_q where $\frac{q-1}{M} \le u_i < \frac{q}{M}$. Then, generate u_j and select the

perturbed value of λ_q as follows. Where α is the chosen maximum displacement allowed at any step, set the perturbed value $\tilde{\lambda}_q$ of λ_q to $\tilde{\lambda}_q = \lambda_q + 2\alpha(u_j - .5)$. If the perturbation carries $\tilde{\lambda}_q$ outside the interval $[\gamma_1;\gamma_2]$ the remaining amount is used to reenter the value from the other side of the interval; i.e.

$$\text{if } \tilde{\lambda}_q < \gamma_1 \qquad \tilde{\lambda}_q = \gamma_2 - \gamma_1 + \lambda_q + 2\alpha(u_j - .5) \tag{3.36}$$

$$\text{if } \tilde{\lambda}_q > \gamma_2 \qquad \tilde{\lambda}_q = \lambda_q + 2\alpha(u_j - .5) - \gamma_2 + \gamma_1$$

If $\Delta E \leq 0$ accept the perturbed configuration. If $\Delta E > 0$ generate another uniform pseudo-random number u_k and accept the perturbed configuration only if $u_k < \exp(\frac{-\Delta E}{T})$.

Metropolis showed that for $\lambda \in \Omega$, where Ω is a finite discrete valued parameter space, a configuration iterated upon by the above algorithm eventually evolves into thermal equilibrium; i.e. the probabability of a configuration with energy E approaches the Boltzmann distribution. This was shown in a much more formal manner by a number of authors through the following method [47]. First, the sequence of configurations generated by the Metropolis is shown to be an irreducible, aperiodic Markov chain. Therefore, a unique stationary distribution exists. Then, the transition probabilities of the Metropolis algorithm and the Boltzmann distribution are shown to satisfy the requirements of the stationary distribution.

The iterative updating of a parameter vector by the Metropolis algorithm can be described by a Markov chain. Let Ω be a discrete parameter space containing M distinct configurations and let ij denote the i^{th} and j^{th} configuration of that space respectively. A Markov chain is characterized by a set of transition probabilities, $\{P_{ij}(\text{k}-1,\text{k})\}$ where $P_{ij}(\text{k}-1,\text{k})$ is the probability that the outcome of the k^{th} iteration is j given the outcome of the $\text{k}-1^{\text{th}}$ iteration is i. If the $P_{ij}(\text{k}-1,\text{k})$'s do not

depend on k, the Markov chain is homogeneous, otherwise it is inhomogeneous. Let $\mathbf{q}_i(\mathrm{k})$ denote the probability of the outcome i at the k^{th} iteration of the algorithm. Then $\mathbf{q}_i(\mathrm{k})$ satisfies the equation

$$\mathbf{q}_i(\mathrm{k}) = \sum_{\mathrm{j}\varepsilon\Omega} \mathbf{q}_j(\mathrm{k}-1)\, P_{ij}(\mathrm{k}-1,\mathrm{k}) \ . \tag{3.37}$$

The stationary distribution of a Markov chain is defined as the unique vector of probabilities $\mathbf{q}$ where

$$\mathbf{q}_j = \lim_{\mathrm{k}\to\infty} P_{ij}(0,\mathrm{k}) \tag{3.38}$$

holds for any i. A stationary distribution of a finite homogeneous Markov chain exists if the chain is irreducible and aperiodic. A Markov chain is irreducible if and only if for all configuration pairs ij there exists a finite n such that $P_{ij}(\mathrm{k},\mathrm{k+n}) > 0$. A sufficient condition for a Markov chain to be aperiodic is that $P_{ii}(\mathrm{k}-1,\mathrm{k}) > 0$ for all i.

The transition probabilities of the Metropolis algorithm can be characterized by the sets of probabilities G_{ij} and $A_{ij}(T)$ where G_{ij} denotes the probability of configuration j being generated as the perturbed configuration given i is the present configuration and $A_{ij}(T)$ denotes the probability of accepting j given the present configuration i, a function of T. Where each element of λ can take M different values, $G_{ij} = \frac{1}{M}$ for configurations j which differ from configuration i by one element; and $G_{ij} = 0$ otherwise. Clearly, $G_{ij} = G_{ji}$. The acceptance probability $A_{ij}(T)$ is equal to $\min\left[1,\exp\left(\frac{E_j - E_i}{T}\right)\right]$

Since $A_{ij}(T) > 0$ for non-zero T and all ij pairs, the Markov chain represented by the Metropolis algorithm is irreducible if the Markov chain

represented by the generation mechanism G_{ij} is irreducible. This holds true for any prescribed order of visiting elements of λ as the perturbed element may take all possible values with some non-zero probability. Therefore the Markov chain is irreducible. Since there is a non-zero probability that the perturbed configuration will equal the present configuration, $P_{ii}(k-1,k) > 0$ for all k , the Markov chain is also aperiodic so that a stationary distribution exists for the Metropolis algorithm.

The stationary distribution $\mathbf{q}$ is uniquely determined by the equations

$$q_i > 0 , \tag{3.39}$$

$$\sum_i q_i = 1 , \tag{3.40}$$

$$q_i = \sum_j q_j P_{ji}(k-1,k) . \tag{3.41}$$

It remains to show that the q_i's defined by the Boltmann distribution satisfy the above equations with the $P_{ji}(k-1,k)$'s defined by the Metropolis algorithm. Let $q_i = \frac{1}{Z}\exp(-E_i/T)$, $Z = \sum_j \exp(-E_j/T)$, and $P_{ji} = G_{ji}A_{ji}(T)$ where $A_{ji}(T) = \min\left[1,\exp(\frac{E_j-E_i}{T})\right]$. Then,

$$\sum_j q_j P_{ji} = \sum_{j\neq i:E_i<E_j} q_j G_{ji} A_{ji}(T) + \sum_{j\neq i:E_i\geq E_j} q_j G_{ji} A_{ji}(T) + q_i P_{ii} \tag{3.42}$$

If $E_i \geq E_j$, then $q_j A_{ji} = q_i$. In addition, $G_{ji} = G_{ij}$ and

$$q_i P_{ii} = q_i \left\{1 - \sum_{j\neq i:E_i<E_j} q_j G_{ij} A_{ji}(T) + \sum_{j\neq i:E_i\geq E_j} q_j G_{ij} A_{ji}(T)\right\} \tag{3.43}$$

Substituting into equation (3.40) results in $\sum_j q_j P_{ji} = q_i$.

The Gibbs sampler [22] is a modification of the Metropolis algorithm. Where the Metropolis algorithm consists of the two steps,

(1) Randomly visit a site λ_i and choose a perturbed site value $\lambda_i = \lambda_i + \Delta\lambda_i$.

(2) Compute the change in energy $\Delta E = E(\lambda+\Delta\lambda) - E(\lambda)$ and accept the perturbed value for the new site value with probability Prob(acceptance) = $\min[1,\exp(\frac{-\Delta E}{T})]$

the Gibbs sampler has the form:

(1) Randomly visit a site λ_i and choose a new site value $\lambda_i = \Delta\lambda_i$ from the distribution $f(\lambda_i \mid \lambda_j : j \neq i)$.

In the later stages of the Metropolis algorithm, a significant amount of computation time is used generating perturbed configurations of low probability which are then rejected. The Gibbs sampler does not suffer from this drawback.

Using the Metropolis algorithm or the Gibbs sampler, we can generate sample realizations from a Gibbs distribution. The images in figures 3.3, 3.4, 3.5, 3.6 and 3.7 were generated using 200 iterations of a Gibbs sampler. These images display the kinds of qualities to which a Gibbs distribution places increased probability. The potential functions defined on cliques containing a single pixel were set to zero. Figure 3.3 has four grey levels on a 128x128 image with a 1st order neighborhood and potential function $V_1(\lambda_i,\lambda_j)$ on nearest neighbors λ_i and λ_j

$$V_1(\lambda_i,\lambda_j) = \begin{Bmatrix} 0.00 & \text{if } \lambda_i = \lambda_j \\ 0.75 & \text{if } \lambda_i \neq \lambda_j \end{Bmatrix} \tag{3.44}$$

Figure 3.4 has four grey levels on a 128x128 image with a 2nd order neighborhood and potential function

$$V_1(\lambda_i,\lambda_j) = \left\{ \begin{array}{ll} 0.00 & \text{if } \lambda_i = \lambda_j \\ 0.75 & \text{if } \lambda_i \neq \lambda_j \text{ and } (\lambda_i\ ,\ \lambda_j) \text{ are nearest neighbors} \\ 0.75 & \text{if } \lambda_i \neq \lambda_j \text{ and } (\lambda_i\ ,\ \lambda_j) \text{ are diagonal neighbors} \end{array} \right\} \quad (3.45)$$

Figure 3.4 shows the increased clustering of pixels with a 2nd order instead of a 1st order neighborhood. Figure 3.5 and 3.6 show larger, 256x256, configurations with 8 instead of 4 grey levels. Figure 3.6 also shows the increased clustering of a 2nd order neighborhood. Figure 3.7 shows a typical clustering for a 1st order neighborhood with the potential function V_1

$$V_1(\lambda_i,\lambda_j) = \left\{ \begin{array}{ll} 0.00 & \text{if } \lambda_i = \lambda_j \\ 1.35 & \text{if } \lambda_i \neq \lambda_j \end{array} \right\}. \quad (3.46)$$

Increasing the penalty for neighboring pixels which are not of the same pixel value brings out clustering even with a 1st order neighborhood.

3.4.2.2 Simulated Annealing

Simulated annealing [43],[47] was introduced as a global optimization method for combinatorial problems by S. Kirkpatrick in 1982. This algorithm consists of the Metropolis algorithm, implemented not at a constant temperature T, but with a gradual lowering of the temperature. The goal is to slowly freeze the configuration at the global maximum of the stationary distribution. As the temperature is lowered, the maxima of the stationary distribution are accentuated so that configurations of lower probability are selected less and less frequently until the probability distribution is concentrated about its global maximum.

The rate of decrease of the temperature is termed the cooling schedule. Let n be an integer such that a global minimum configuration can be reached from any other configuration in no more than n transitions. Furthur, let

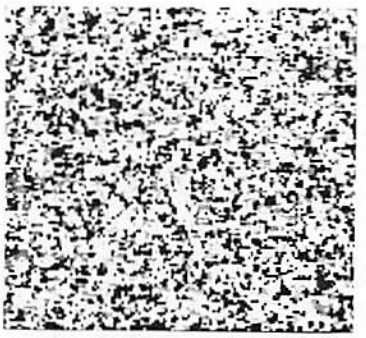

Figure 3.3

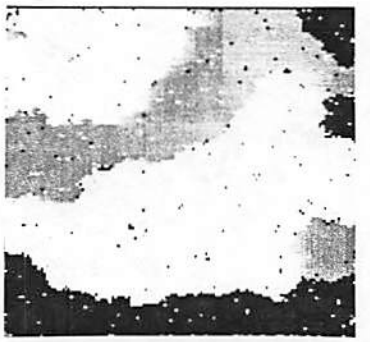

Figure 3.4

128x128 sample configurations from Markov random fields with four grey levels. Figure 3.3 is based on a 1st order neighborhood with the potential function in equation (3.44) and Figure 3.4 is based on a 2nd order neighborhood with the potential function in equation (3.45).

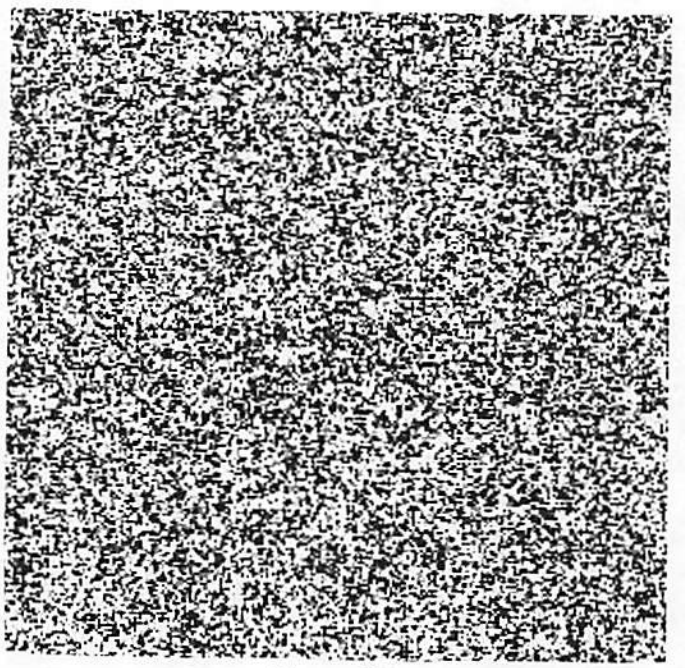

Figure 3.5

A 256x256 sample configuration from a Markov random field based on a 1st order neighborhood with the potential function in equation (3.44) and eight grey levels.

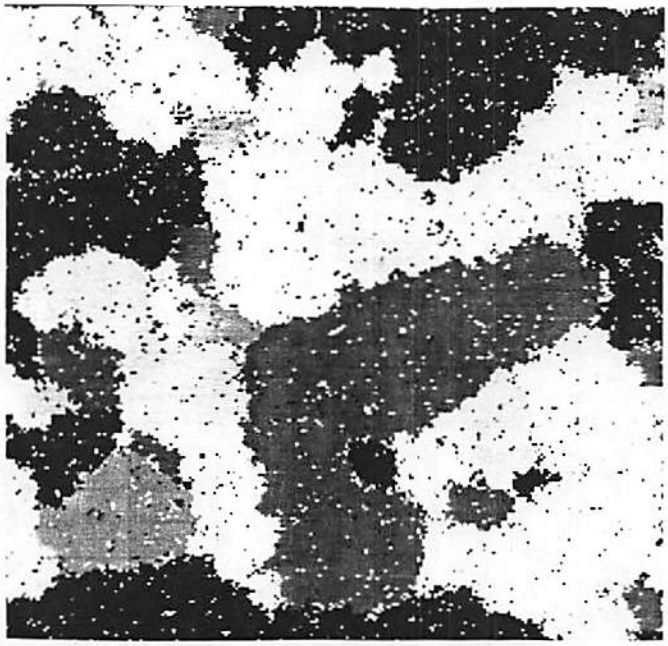

Figure 3.6
A sample Markov random field based on a 2nd order neighborhood with the potential function in equation (3.45) and 8 grey levels.

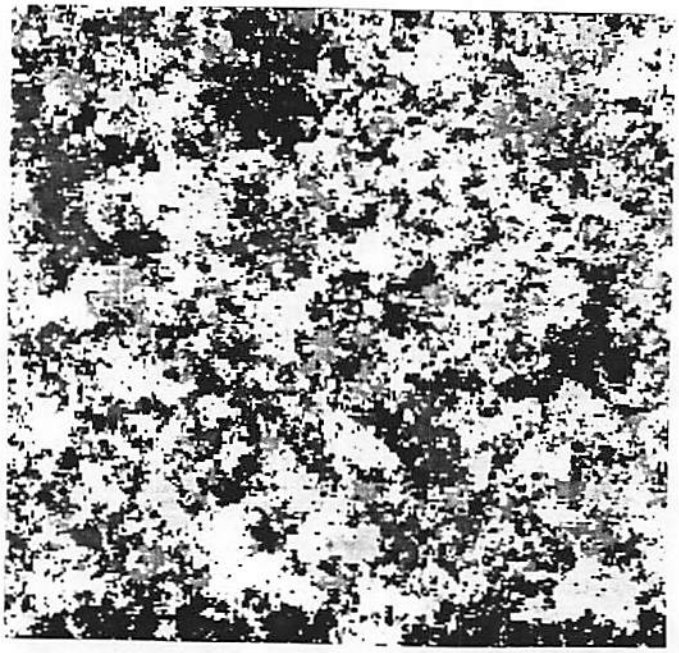

Figure 3.7
A sample Markov random field based on a 1st order neighborhood with the potential function in equation (3.46) and 8 grey levels.

$\Delta = \max_{i,j}\{E_j - E_i \mid E_j > E_i\}$. Where k is the iteration number of the algorithm, a cooling schedule for which $T > \frac{n\Delta}{\log(k)}$ is sufficiently slow to ensure convergence to the globally minimum configuration. The algorithm is terminated for some small value for T at which virtually no perturbations are accepted. The final configuration is taken as the solution to the problem.

Simulated annealing algorithm is easily applied to any combinatorial optimization problem in which the parameters may take values from a finite discrete set. Geman and Geman [22] applied this algorithm to image processing. If the parameters to be estimated are allowed to take a continuous range of values, simulated annealing can no longer be applied. In some applications, simulated annealing may require dramatically more computation than deterministic algorithms. The benefit of escaping local minima does not come without a computational cost. The applications of simulated annealing to image restoration and reconstruction which appear in the literature generally do not include a blurring function in their simulations. This reduces not only the computational cost of each iteration, but also the number of iterations required for convergence. As a general rule, any restoration or reconstruction algorithm requires more iterations to converge as the blurring function becomes more extensive.

When the computational cost is a factor, the time required to perform the restoration or reconstruction is a factor, or when the dimension of the problem is large enough, the cost of using simulated annealing may be prohibitive. In addition, for many problems, the quality of estimates at local maxima may be sufficiently comparable to those at the global maximum. In these situations, a deterministic algorithm may provide a completely acceptable alternative. Having examined the stochastic algorithms and the required computation, in the next chapters we focus on

developing deterministic algorithms for solving probabilistically posed restoration and reconstruction problems.

Chapter 4

Statistical Stopping Criterion for Iterative Algorithms

4.1 Introduction

Many investigators have observed [34],[41],[48],[54],[63],[77],[79],[80] that least squares (LS) and maximum likelihood (ML) restorations and reconstructions of noisy, blurred images are excessively non-smooth unless the SNR is very high. However, when these approaches are implemented iteratively, the reconstructions do not develop this quality until after the early stages of the iterative process. Therefore, when iterative algorithms are used, the iterations are typically terminated well before convergence, since this produces the best 'looking' results. The need for a strategy with a formalized stopping criteria and a justification for terminating iterative algorithms early has not been sufficiently addressed.

In the signal processing literature, Trussel [79] proposed an alternative convergence criterion for iterative least squares based on the linear Gaussian model $\mathbf{y} = \mathbf{P}\boldsymbol{\lambda} + \mathbf{n}$, where $\mathbf{y}$ is the degraded image, $\boldsymbol{\lambda}$ is the undegraded image, $\mathbf{P}$ is a blurring function, and $\mathbf{n}$ is white noise. Instead of the favored convergence criterion $\| \mathbf{y} - \mathbf{P}\hat{\boldsymbol{\lambda}} \|_2^2 < \varepsilon$, where ε is chosen as some small value, Trussell examined the following sum of normalized squared errors. If the variances of the noise components are known,

$$\chi^2 = \sum_i^N \frac{[\lambda_i - (\mathbf{P}\hat{\boldsymbol{\lambda}})_i]^2}{\sigma_i^2} \tag{4.1}$$

has a chi-squared distribution with N degrees of freedom. Trussell argued that convergence of iterative least squares algorithms should be considered to occur

when the statistic χ^2 lies within a 95% confidence interval. He further argued that the Landweber least squares algorithm [73] would at some iteration generate an estimate satisfying the criterion. The motivation for this approach was that the restored image looked better if the iterations were terminated earlier than dictated by conventional numerical convergence.

A similar proposal was made by Veklerov and Llacer [80] for the Poisson data model (section 2.2). Using the estimate of the undegraded image and the data, a set of pseudo-random numbers can be generated which would be i.i.d. uniform (0,1) if the image estimate were the true undegraded image. A variety of statistical tests can then used to determine if these random numbers sufficiently follow a uniform (0,1) distribution. The pseudo-random numbers were based on the discrete equivalent to the fact that if x is a continuous random variable with distribution function $F(\cdot)$, then $F(x)$ is uniformly (0,1) distributed. Let $\overline{y}_i$ equal the i^{th} element of matrix-vector product $\mathbf{P}\hat{\boldsymbol{\lambda}}$. Given the estimated mean $(\mathbf{P}\hat{\boldsymbol{\lambda}})_i$ of each data value y_i at a particular iteration, the probability sums

$$t_1 = \sum_{n=0}^{y_i-1} \frac{(\overline{y}_i)^n \exp[-\overline{y}_i]}{n!} \tag{4.2}$$

$$t_2 = \sum_{n=0}^{y_i} \frac{(\overline{y}_i)^n \exp[-\overline{y}_i]}{n!}$$

are evaluated. Since t_1 and t_2 are functions of the random data, a pseudo-random uniform number between t_1 and t_2 is uniformly distributed (0,1). The U(0,1) distribution is divided into intervals and Pearsons chi-squared test is applied to the number of pseudo-random variables falling in each interval [50]. This approach suffers from two shortcomings: 1) There is no indication that a standard ML

iterative algorithm will arrive at an estimate satisfying the statistical criterion. In fact, as the Gaussian distribution is divided into more intervals, it can be argued that the probability the criterion will be satisfied approaches zero. 2) the number of intervals into which the uniform distribution function is divided is arbitrary and will affect the performance of the approach.

Our presentation [34], which follows, differs from Trussells [79] and Veklerov and Llacers [80] in that we recognize the role that the initial estimate of the iterative algorithm plays. We present the approach as a path-dependent regularization of the optimization criterion based upon the initial estimate. Furthur, we show that, with probability approximately equal to 1, any iterative ML algrithm will generate an estimate satisfying our stopping criterion. We address the Poisson data model, equation (2.8), and we compare the performance of this approach using four different statistics.

4.2 A Perspective on Early Termination of Iterative Algorithms

When applying iterative LS or ML approaches to restoration or reconstruction of blurred and noisy images, ideally, one would continue iterating until the Kuhn-Tucker conditions are satisfied [52]. Most of these iterative algorithms initially converge quite quickly yet produce very slow convergence in the later iterations. In emission tomography, for example, more than a hundred ML or LS iterations may be required for the Kuhn-Tucker conditions to be satisfied to several digits. However, after far fewer iterations, the reconstructions generally deteriorate by taking on an excessive, non-smooth appearance. From simulations, which are shown later, we see that near this point the squared error between the estimate and the true image begins to increase. The most attractive and accurate restorations and

reconstructions are achieved by halting the ML or LS algorithm relatively early in the optimization process.

On first impression, it would seem incorrect to formulate a solution based on a certain criterion, yet then fail to proceed to that solution. A justification for using a stopping criterion needs some further insight. When a stopping criterion is employed, the initial estimate plays a significant role in the reconstruction. To evaluate this role, we take a closer look at the iterative reconstruction process. Iterative reconstruction algorithms are typically initialized with a uniform, constant estimate. An algorithm then iterates towards the solution in steps of varying size and direction. Given the data and the initial estimate, the path is specified by the particular choice of optimization algorithm. Such a process for a ML algorithm is represented in figure 4.1 where the dotted contours represent contours of constant likelihood. In halting the algorithm prematurely, one accepts some estimate along the optimization path, as at the point marked "X" in figure 4.1.

A number of investigators have noticed that image reconstructions from blurred and noisy observations become increasingly non-smooth in later iterations of ML and LS algorithms. An example of this degradation is shown in the images along the top row of figure 4.2 which show the evolution of an emission image reconstruction from a constant initial source estimate through 240 iterations of the EM algorithm. Starting from a constant initial estimate, the reconstruction of a source from mechanically collimated data retains some of the smoothness of that initial estimate for a number of iterations of a ML algorithm. After some 30 iterations of the EM ML algorithm [64] the reconstruction becomes excessively non-smooth. The degradation of the reconstruction in iterations past 30 can be put into perspective by beginning, not with an initial constant source estimate, but with a

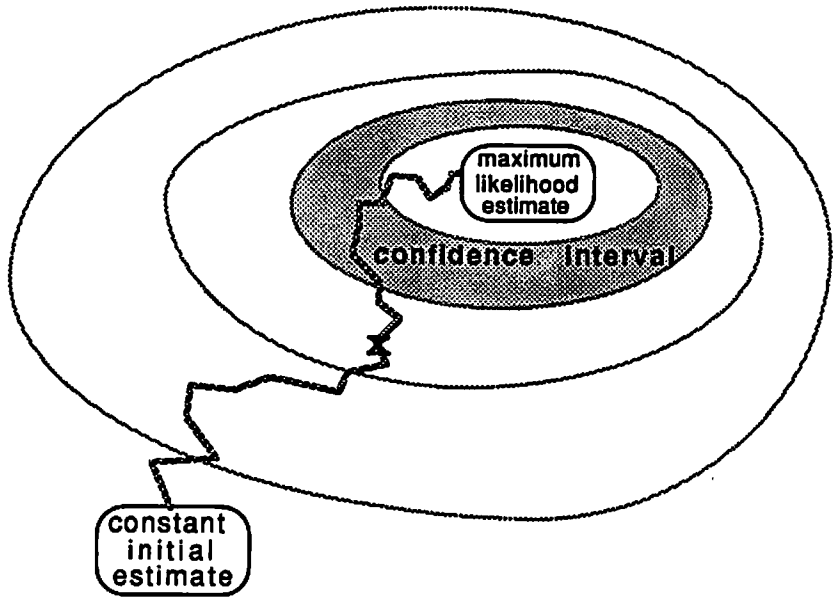

Figure 4.1
An Iterative Optimization Process

pseudo-random estimate generated from a uniform distribution plus a constant. Proceeding from this non-smooth initial estimate, (bottom row of figure 4.2) the reconstruction loses some irregularity as the iterations proceed, and one has no qualms with adhering to the ML approach and continuing the iterations until the ML solution is reached. A degradation is only apparent when observing the sequence of estimates from a smooth initial estimate. The observed degradation is simply ones perception of the convergence process from the smooth initial estimate to the non-smooth maximum likelihood solution. A similar effect can be seen in the case of maximum likelihood restoration of blurred and noisy images. Figure 4.3 shows restorations using a uniformly smooth initial estimate and using a random initial estimate after 4 and 16 iterations of an iterative maximum likelihood algorithm. Strictly speaking, it is inappropriate to speak of a degradation of the

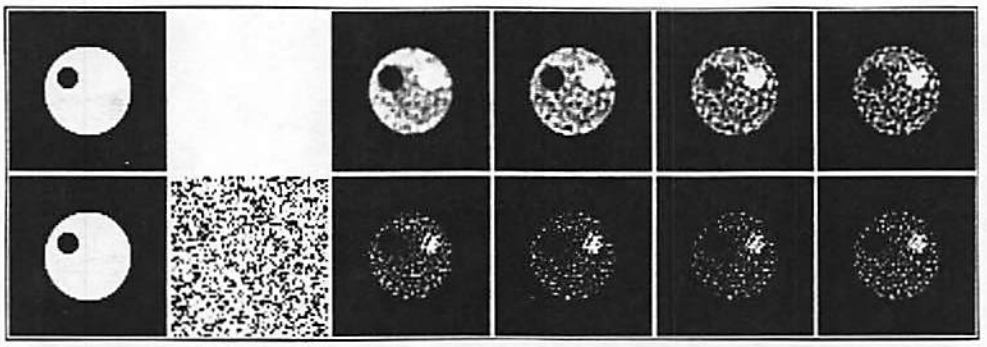

Figure 4.2
ML reconstructions from a parallel-hole collimated system. From left: True images, initial estimates, results from 20, 60, 120, 240 EM iterations.

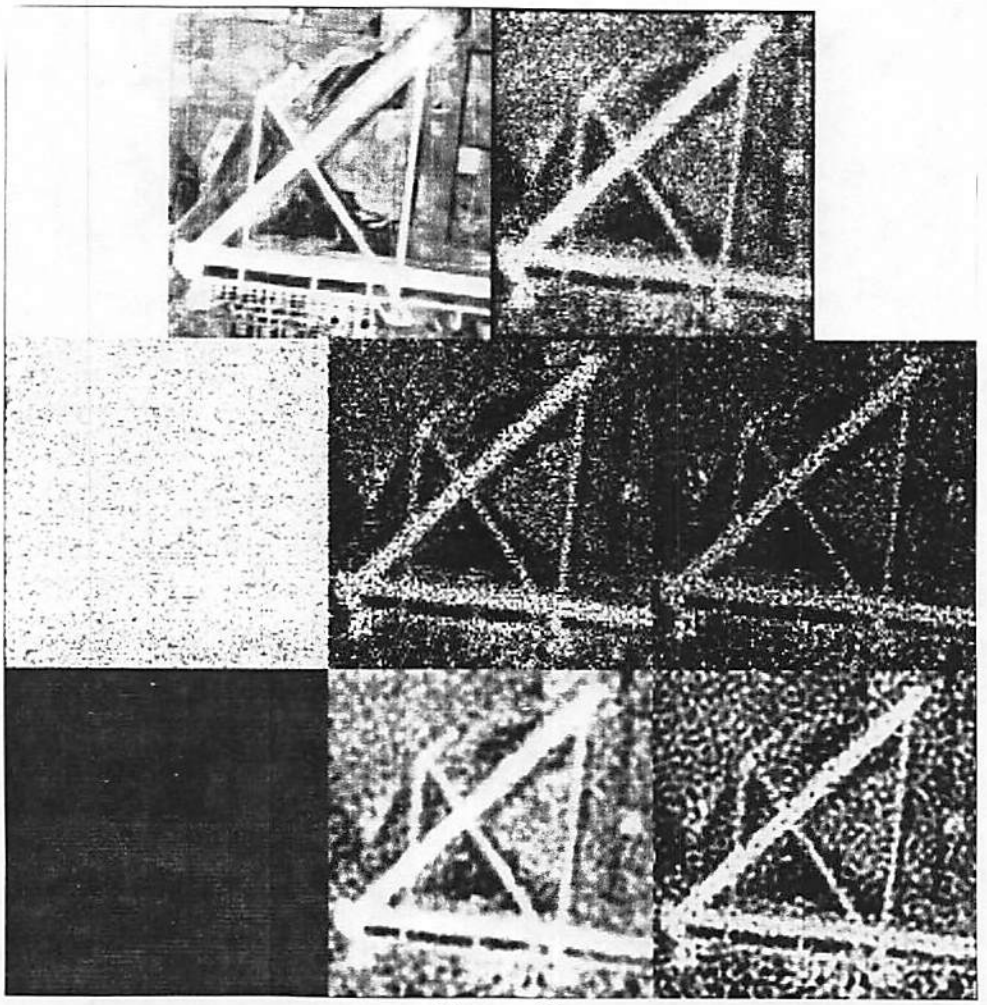

Figure 4.3
From Left: Top row - image, image with 7x7 blur and Poisson noise. Middle and bottom rows - initial estimate, results from 4 and 16 ML iterations.

estimate as the iterations proceed since an estimate satisfying the maximum likelihood restoration criterion is only achieved at convergence. But it is certainly appropriate to discuss the unacceptable non-smooth quality of the maximum likelihood solution in many image restoration and reconstruction applications and view the restoration as a path-dependent process.

Starting from a constant initial estimate λ_c, it is apparant that an iterative ML reconstruction retains some of the smoothness of that initial estimate during the intial iterations. The effect of the constant initial estimate has been to anchor the free end of the optimization path (i.e. the initial point) such that the path passes through smoother estimates than those traversed when the algorithm is initialized with a pseudo-random estimate. In fact, given the assumption that images possess a degree of smoothness, such a path may in general pass through estimates much closer to the true underlying image than the ML estimate. One would therefore like some strategy of effecting a trade-off between the smoothness of the constant initial estimate λ_c and the asymptotic consistency and efficiency of the ML estimate λ_{ML}. Such an approach also provides a justification and the method for applying smoothed versions of iterative algorithms.

For a given set of data, confidence intervals about a statistic define broad, non-convex regions in the undegraded image space which, with high probability, contain the true underlying image. A trade-off between λ_c and λ_{ML} is most aptly defined by a confidence region. The determination of a stopping point can be posed as an hypothesis test. Statistics for which this test performs well for sources within a wide range of complexity are sought. The performances of these statistics are compared.

4.3 An Hypothesis Test Approach

In using an hypothesis test [50] as a stopping criteria, the null hypothesis is that the data y are Poisson with vector of parameters ($\mathbf{P}\hat{\boldsymbol{\lambda}}^k$), where $\hat{\boldsymbol{\lambda}}^k$ is the estimate at the k^{th} iteration. A test statistic S_t having a known density function under a correct null hypothesis is formulated. A critical region, consisting of all values outside some confidence interval $[a,b]$, and a significance level α are chosen such that if the null hypothesis is correct $P(a<S_t<b) = 1-\alpha$. The hypothesis is tested by computing the statistic S_t, a function of $\boldsymbol{\lambda}^k$ and $\mathbf{y}$. If $S_t<a$ or $S_t>b$ the null hypothesis is rejected at the α significance level and an alternate estimate of $\boldsymbol{\lambda}$ is sought. Otherwise, the estimate $\hat{\boldsymbol{\lambda}}^k$ is accepted and the algorithm terminated.

4.4 Suitable Statistics

We examine two approaches to developing test statistics for use as a stopping criterion.

(1) By conditioning the multivariate Poisson distribution function of the data upon the total number of counts, a multinomial distribution function which fits naturally into Pearson's chi-squared statistic is achieved [58].

(2) Given an estimate of the mean of the data, a pseudo-random set of Poisson data $\mathbf{y}^*$ can be generated. Statistics based upon the distribution of the paired differences $y_i - y_i^*$ can be derived under the null hypothesis.

In appendix 4a, a form of Pearson's statistic ${\chi_p}^2$ is formulated for the case of Poisson data whose mean is equal to a blurred version of the uncorrupted image as in chapter 2, equation (2.4). Where S_d is the sum of all data, Pearson showed that as $S_d \rightarrow +\infty$ the distribution of ${\chi_p}^2$ converges to a chi-squared distribution with $M-1$ degrees of freedom. In evaluating ${\chi_p}^2$, if N of the parameters have been

estimated by any asymptotically normal and asymptotically efficient estimator, the distribution of χ_p^2 is asymptotically chi-squared with $M-1-N$ degrees of freedom [46].

χ_p^2 evaluated at $\lambda=\lambda_c$, a uniform constant initial estimate, will invariably result in $\chi_p^2 \gg b$ because λ_c does not adhere at all to the data. Alternatively, if χ_p^2 is evaluated at $\hat{\lambda}_{ML}$ the maximum likelihood solution, invariably $\chi_p^2 \ll a$. The latter is due to the fact that N parameters have been estimated by the asymptotically normal and asymptotically efficient ML estimator, χ_p^2 has $M-1-N$ degrees of freedom, and is expected to have much lower value. However, the use of Pearson's chi-squared statistic in effecting a trade-off between the influence of λ_c and the asymptotic efficiency and consistency of $\hat{\lambda}_{ML}$ is to determine the first iteration at which the estimate agrees with the data without satisfying the constraints imposed by the Kuhn-Tucker conditions. Therefore, in this context, a two-sided critical region about a confidence interval for a chi-squared variable of $M-1$ degrees of freedom is the region of interest. It is clear that χ_p^2 evaluated along a ML optimization path from λ_c to $\hat{\lambda}_{ML}$ will decrease in value from above an interval expected of a chi-squared variable of $M-1$ degrees of freedom down towards some interval expected of a chi-squared variable of $M-1-N$ degrees of freedom. Provided that N is at least a significant fraction of M, we can therefore conclude that Pearson's chi-squared statistic will with probability approximately equal to one indicate a sufficient fit of the estimate to the data at some point in the optimization process. The optimization process using the EM ML algorithm [64] and Pearson's chi-squared statistic as a stopping criterion has the form shown in figure 4.1, where the confidence interval based on $N-1$ degrees of freedom surrounds but does not include the maximum likelihood estimate.

The three remaining statistics to be developed are based on a user generated set of independent Poisson pseudo-random variables. Given the observation vector y, the matrix of probabilities **P** , and the estimate vector λ^k at iteration k; we form the vector of data means $\mathbf{P}\lambda^k$. We then generate a set of independent pseudo-random Poisson variables $\hat{y}$ with mean $\mathbf{P}\lambda^k$ and form the variables

$$\xi_j = \frac{y_j - \hat{y}_j}{\sqrt{2(\mathbf{P}\lambda)_j}} \tag{4.3}$$

Statistics based on these variables are of interest in that under the null hypothesis they represent the normalized differences between the data and an alternate pseudo-random realization of the data. We first examine the ordinary sign statistic which we label η_s. We note that under the null hypothesis, y_j and $\hat{y}_j$ have the same mean and therefore $\text{Prob}(\xi_j < 0) = \text{Prob}(\xi_j > 0)$ so that the ξ_j are symmetrically distributed about zero. If form the variables

$$x_j = \begin{cases} 0 & \text{if } \xi_j > 0 \\ 1 & \text{if } \xi_j < 0 \\ \frac{1}{2} & \text{if } \xi_j = 0 \end{cases} \tag{4.4}$$

and discard the K x_i's which equal $\frac{1}{2}$, the remaining variables x_j are independent Bernoulli random variables [56]. The sum $X_s = \sum_{j=1}^{M-K} x_j$ has a binomial distribution with probability of success $\frac{1}{2}$ and is known as the ordinary sign statistic [56]. For $M-K$ large, $\eta_s = \frac{2X_s-(M-K)+1}{\sqrt{M-K}}$ is therefore approximately standard unit normal. The advantages of using the ordinary sign statistic are that it is easily computed and it can be used where the variables of interest do not necessarily satisfy any

assumptions other than that each has a symmetric probability function.

We secondly examine Wilcoxon's signed rank statistic T^- which gives more weight to variables ξ_i with large magnitudes than to those with small magnitudes. If the variables $(\xi_1 \cdots \xi_M)$ are ranked according to magnitude, where r_i is the rank, 1 through M, of ξ_i and x_i are defined as above, Wilcoxon's signed rank statistic is

$$T^- = \sum_{i=1}^{M} r_i x_i \tag{4.5}$$

For large M, T^- is asymptotically normal with mean and variance [56]

$$\begin{aligned} E\{T^-\} &= \frac{1}{4}M(M+1) \\ \text{var}\{T^-\} &= \frac{1}{24}M(M+1)(2M+1) \end{aligned} \tag{4.6}$$

The statistic $T_N^- = \frac{T^- - E\{T^-\}}{\sqrt{\text{var}\{T^-\}}}$ is therefore asymptotically standard unit normal.

To derive the final statistic we note that the probability function of ξ_j in equation (4.3) can be derived as in appendix 4b. Under the hypothesis that λ^k is the known true vector of parameters, the final statistic, $\psi_s^2 = \sum_m^M \xi_m^2$, is asymptotically chi-squared with M degrees of freedom (appendix 4b). This is dependent upon the mean number of counts in each term $(\mathbf{P}\lambda)_j \to +\infty$. An asymptotic dependence on $N_d \to +\infty$ such as Pearson's χ_p^2 statistic and the ordinary sign statistic T^- is more desireable than a dependence on $(\mathbf{P}\lambda)_j \to +\infty$ such as ψ_s^2 since N_d, the total number of detected counts, is much larger than the means $(\mathbf{P}\lambda)_i$.

The cumulative distribution function of ξ_j (appendix 4b), which is a step-wise approximation to the standard normal, can be plotted as in figure 4.3. This is done

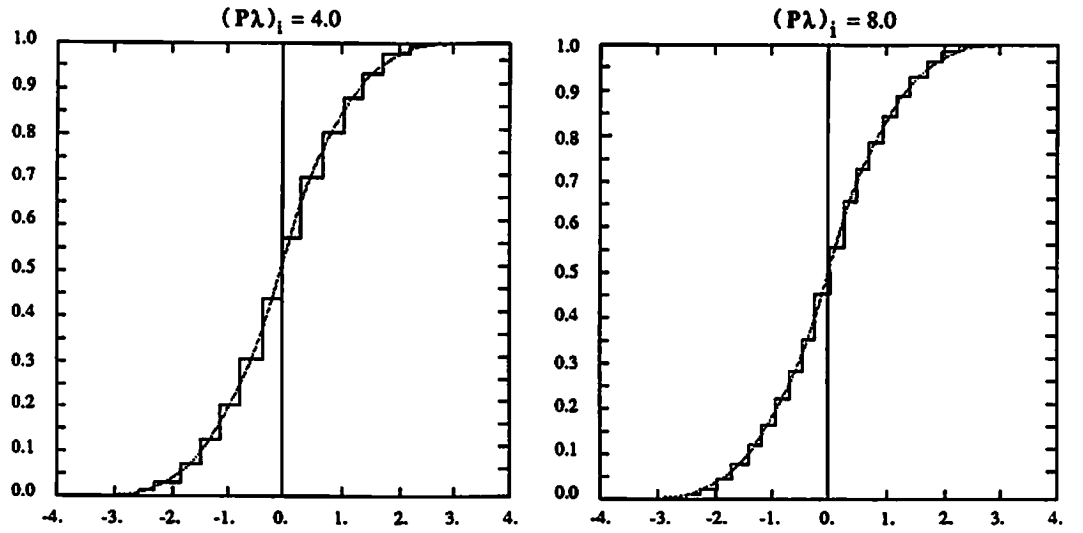

Figure 4.4
Solid line - cumulative distribution function of ξ in equation (4.2). Left graph - $(P\lambda)_j = 4.0$. Right graph- $(P\lambda)_j = 8.0$. Dotted line - cumulative distribution function of standard normal variable.

by using tabled values of $e^{-x}I_0[x]$ and $e^{-x}I_1[x]$ in conjunction with the fact that $e^{-x}I_{n+1}[x] = e^{-x}I_{n-1}[x] - \frac{2n\,e^{-x}}{x}I_n[x]$ [81] where $I_n[\cdot]$ is the modified Bessel function of the first kind order n. A larger mean $(P\lambda^k)_j$ gives a smaller step size and improves the approximation.

The variables ξ_j are independent as are sums having no variables in common. Where necessary, the M terms can be pooled [50],[56] into M_p terms to ensure, for instance, $(P\lambda)_j \geq 12$ for each ξ_j before χ_s^2 is calculated. At each iteration of an iterative LS or ML algorithm, the changing estimate $\hat{\lambda}^k$ and the pooling of means may give a statistic with a slightly different number of degrees of freedom. Therefore the the critical region adapts somewhat to the localization of intensity in the estimate.

estimate.

4.5 Results and Conclusions

In order to evaluate the performance of the hypothesis test approach to halting iterative ML algorithms, a parallel-hole collimated single photon emission imaging system with a 25 cm diameter gamma camera was simulated. Perfect collimation between 0.4 cm thick horizontal slabs was assumed allowing single images to be reconstructed. The mean of the data was generated as a discrete approximation to the weighted integral over each strip of the mean source activity, the weighting at a given source location being proportional to the solid angle subtended by the pixel of the gamma camera. This mean was used in generating Poisson random variables representing the number of recorded counts within each pixel of the gamma camera. The mean activity of the source phantom thus translates into the mean number of counts of the data.

Four different simulated source phantoms were used in this study (figure 4.4). Each of the four phantom was reconstructed from two different mean levels of activity, those generating a total mean number of counts of 100,000 and 200,000.

For performance measures, two measures of the difference between the reconstructed source and the true source are examined, average absolute error (L_1) and average squared error (L_2). It was found for the simulation presented, that 32 iterations of the EM ML algorithm achieved the lowest average squared error and 45 iterations achieved the lowest average absolute error. Of course, these numbers were found after the simulations were completed and could not be predicted beforehand by any method. However, they serve as a comparison between an optimal fixed number of iterations and the statistical approach presented in this chapter. The

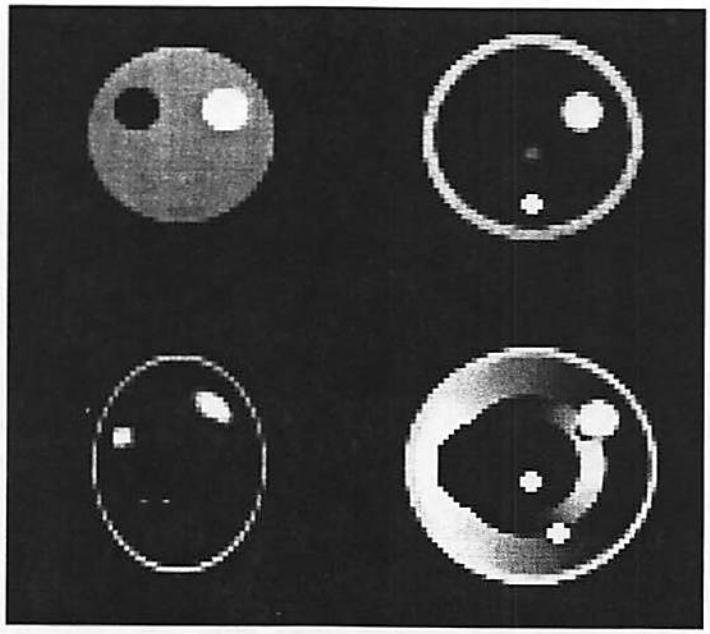

Figure 4.5
Phantoms 1-4 used to evaluate the stopping criteria

results using a fixed number of iterations, 32 and 45, are compared to those obtained using the statistical criterion.

The hypothesis test approach to determining a suitable stopping point in the iterative ML reconstruction proceeds as in figure 4.5. Figure 4.5 shows a typical plot of the value of statistic χ_p^2, a 95% confidence interval, absolute image estimate error, and squared image error for 100 iterations of the EM ML algorithm. An indication to halt iterations is given by each statistic as it crosses the mode of the distribution function. Such plots were made for each of the four statistics and each of the four phantoms in figure 4.4. In reconstructions of the four simulated phantoms from both 100,000 and 200,000 mean counts, the iteration halting points generated by the four statistics along with the number of iterations resulting in the minimum

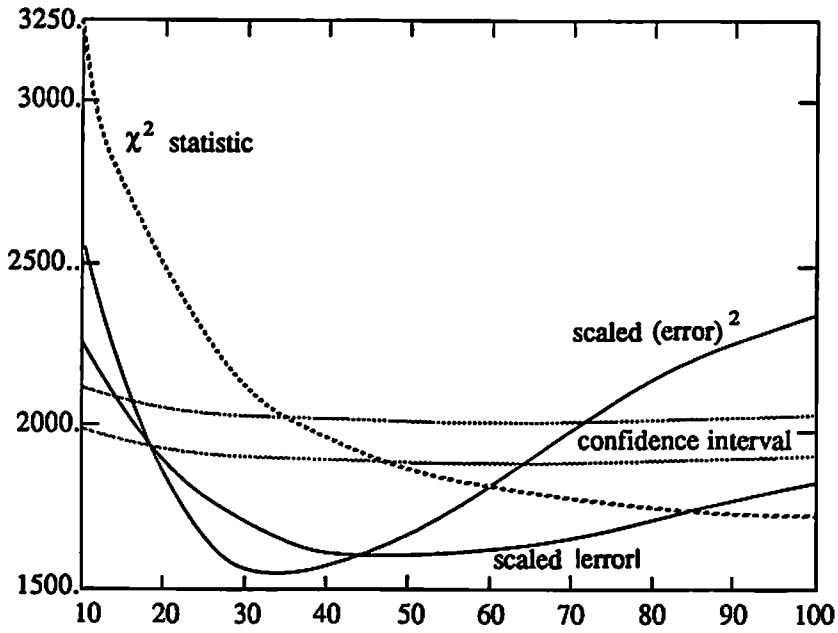

Figure 4.6
χ_p^2 statistic, squared error, absolute error, and a confidence interval vs. 100 iterations of the EM algorithm.

average squared error and average absolute error are listed in Table 4.1. The likelihoods at the iteration stopping points determined by the four test statistics are compared in table 4.2. This demonstrates that the likelihoods at iteration stopping points determined by the statistical criterion are not significantly lower than those of maximum likelihood. The average L_2 reconstruction error of the reconstruction at the iteration indicated point determined by the hypothesis test is compared to the minimum L_1 reconstruction error achieved along the optimization path to the ML solution λ_{ML} in table 4.3. Table 4.4 compares the absolute error between the reconstructed images and the true undegraded image for the statistical stopping points as indicated.

	χ_p^2	η_s	T_N^-	ψ_s^2	min $(err)^2$	min |err|
phantom 1a	14	3	9	13	15	16
phantom 2a	30	24	56	26	35	60
phantom 3a	31	19	17	29	35	45
phantom 4a	19	1	13	21	25	28
phantom 1b	18	7	15	15	19	19
phantom 2b	47	44	68	35	42	100
phantom 3b	37	33	39	38	52	60
phantom 4b	26	10	23	22	34	37

Table 4.1
A tabled comparison of *number of iterations* indicated by the four test statistics with the number of iterations yielding minimum true squared error and minimum true absolute error in EM algorithm reconstructions of 8 phantom reconstructions.

	χ_p^2	η_s	T_N^-	ψ_s^2	32 iters	45 iters	min $(err)^2$	min |err|	max like
phantom 1a	290180.	271849.	289337.	290116.	290426.	290426.	290229.	290262.	290546.
phantom 2a	274456.	274343	274603.	274391.	274481.	274481.	274509.	274611.	274693.
phantom 3a	292559.	292347.	292259.	292538.	292566.	292566.	292587.	292634.	292748.
phantom 4a	261430.	242488.	260932.	261496.	261656.	261656.	261579.	261619.	261751.
phantom 1b	721515.	716498.	721356.	721356.	721708.	721708.	721547.	721547.	721876.
phantom 2b	690353.	690332.	690442.	690226.	690169.	690169.	690314.	690492.	690534.
phantom 3b	721399.	721357.	721419.	721409.	721341.	721341.	721497.	721526.	721630.
phantom 4b	661330.	658622.	661238.	661198.	661430.	661430.	661453.	661481.	661688.

Table 4.2
A comparison of the *likelihood* at the iteration stopping points indicated by the four test statistics with the likelihood after the number of iterations yielding minimum true squared error and minimum true absolute error in EM algorithm reconstructions of 8 phantom reconstructions.

	χ_p^2	η_s	T_N^-	ψ_s^2	32 iters	45 iters	min $(err)^2$	min lerrl	max like
phantom 1a	111.8	1095.3	154.9	114.3	183.5	269.3	110.9	111.3	1024.4
phantom 2a	504.2	538.9	537.1	522.3	500.3	511.0	498.4	547.7	898.6
phantom 3a	525.2	614.1	652.1	530.3	523.5	535.9	521.3	535.9	1148.5
phantom 4a	230.2	2259.8	287.3	223.4	229.8	275.2	219.1	221.6	755.6
phantom 1b	377.8	886.7	395.0	395.0	478.3	648.4	377.6	377.6	2408.6
phantom 2b	1641.7	1638.8	1674.5	1652.0	1673.4	1639.6	1638.1	1728.0	1940.5
phantom 3b	1623.8	1688.7	1600.2	1611.3	1709.4	1556.3	1541.7	1556.1	2487.7
phantom 4b	694.7	1418.2	730.5	746.9	665.5	694.2	664.0	667.0	1556.7

Table 4.3

A comparison of the *average true squared pixel error* at the iteration stopping points indicated by the four test statistics with that after the number of iterations yielding minimum true squared error and minimum true absolute error in EM algorithm reconstructions of 8 phantom reconstructions.

	χ_p^2	η_s	T_N^-	ψ_s^2	32 iters	45 iters	min $(err)^2$	min lerrl	max like
phantom 1a	4.505	24.243	5.708	4.579	5.455	6.433	4.470	4.464	12.032
phantom 2a	9.430	10.202	8.678	9.876	9.278	8.798	9.105	8.670	10.065
phantom 3a	8.651	9.733	10.103	8.756	8.609	8.380	8.508	8.380	10.913
phantom 4a	7.874	41.624	9.361	7.655	7.452	7.931	7.444	7.414	11.999
phantom 1b	7.930	15.187	8.147	8.147	8.719	9.952	7.915	7.915	18.334
phantom 2b	15.863	16.026	15.065	16.757	17.159	15.970	16.150	14.545	14.551
phantom 3b	14.841	15.298	14.657	14.746	15.436	14.286	14.068	13.975	16.513
phantom 4b	13.283	22.155	13.717	13.920	12.856	12.880	12.799	12.769	18.039

Table 4.4

A comparison of the *average true absolute pixel error* at the iteration stopping points indicated by the four test statistics and at the number of iterations yielding minimum true squared error and minimum true absolute error in EM algorithm reconstructions of 8 phantom reconstructions.

At present, the LS and ML approaches to image restoration and reconstruction are generally applied by iterating until convergence or arbitrarily stopping the algorithm before convergence. The use of confidence intervals in an hypothesis test approach has been shown to provide a means of determining suitable stopping points for iterative ML algorithms. This method has been motivated by the observation that iterative reconstruction algorithms initiated with a uniformly smooth image may pass through estimates which are closer to the true image than the ML estimate. In terms of L_1 and L_2 reconstruction error, this method performed better than both simple maximum likelihood and iterating an optimal fixed number of iterations as determined after all reconstructions had been performed. This method provides a simple alternative to regularization and Bayesian approaches, is less complicated to implement, and requires fewer computations.

4.6 Appendix 4a

Pearsons chi-squared statistic can be applied to estimates of the parameters of a multivariate Poisson distribution in the following manner. For known λ , the probability distribution of photon emission data y conditioned upon $\sum_{m=1}^{M} y_m = N_d$ is given by

$$P(\mathbf{y} \mid \sum_m y_m = N_d) = \frac{P(\mathbf{y})}{P(\sum_m y_m = N_d)} = \frac{N_d!}{y_1! \cdots y_M!} (\pi_1)^{y_1} \cdots (\pi_M)^{y_M} \tag{4.7}$$

where $\pi_k = \frac{(P\lambda)_k}{\sum_m (P\lambda)_m}$. This is a multinomial distribution with probability π_i in the i^{th} of M classes. Pearson's chi-squared statistic ${\chi_p}^2$ is formed from this multinomial sample as

$$\chi_p^2 = \sum_{m=1}^{M} \frac{[y_m - N_d \pi_m]^2}{N_d \pi_m} \tag{4.8}$$

When this statistic is applied to estimates formed using the EM ML algorithm, it has a simplified form. It is easily shown that every iteration of the EM algorithm insures $\sum_m y_m = \sum_m (P\lambda)_m$ [41]. For any estimate λ^k generated by this algorithm, χ_p^2 reduces to

$$\chi_p^2 = \sum_m \frac{[y_m - (P\lambda)_m]}{(P\lambda)_m} \tag{4.9}$$

the sum of the normalized squared forward projection errors.

4.7 Appendix 4b

The statistic derived in this appendix $\psi_s^2 = \sum_m \xi_m^2$ where ξ_m^2 is defined in equation (4.2) will be shown to be asymptotically chi-squared with M degrees of freedom, as follows. Under the hypothesis that λ^k is the known true vector of parameters, the probability function of ξ_j can be derived as follows.

Under the null hypothesis both $E\{y_j\} = (P\lambda)_j$ and y_j and $\hat{y}_j$ are independent, whereby for n≥0

$$\text{Prob}\left(\xi_j = \frac{n}{\sqrt{2(P\lambda)_j}}\right) = \text{Prob}(y_j - \hat{y}_j = n) = \sum_{k=0}^{+\infty} \text{Prob}(y_j = n+k)\ \text{Prob}(\hat{y}_j = k) \tag{4.10}$$

$$= \sum_{k=0}^{+\infty} \frac{(P\lambda)_j^{n+2k} e^{-2(P\lambda)_j}}{(n+k)!\,(k)!} = e^{-2(P\lambda)_j}\, I_n[2(P\lambda)_j] \qquad n \geq 0$$

where $I_n[\cdot]$ is the modified Bessel function of the first kind and order n. By symmetry,

$$\text{Prob}(\xi_j = \frac{n}{\sqrt{2(P\lambda)_j}}) = e^{-2(P\lambda)_j} I_{|n|}[2(P\lambda)_j] \qquad n=0,\pm1,\pm2,\ldots \quad (4.11)$$

The statistic $\psi_s{}^2 = \sum_m^M \xi_m{}^2$ is easily shown to have mean M and variance $2M + \sum_m \frac{1}{2(P\lambda)_m}$.

Under the null hypothesis the variables ξ_j are also independent. Their joint characteristic function is given by

$$\phi_\xi(t_1,\ldots,t_M) = \prod_{m=1}^{M} E\{e^{it_m\xi_m}\} = \prod_{m=1}^{M} E\{e^{\frac{it_m y_m}{\sqrt{2(P\lambda)_m}}}\} E\{e^{\frac{-it_m \vartheta_m}{\sqrt{2(P\lambda)_m}}}\} \quad (4.12)$$

$$E\{e^{\frac{it_m y_m}{\sqrt{2(P\lambda)_m}}}\} = \sum_{n=0}^{+\infty} e^{-(P\lambda)_m} \frac{\left[(P\lambda)_m e^{\frac{it_m}{\sqrt{2(P\lambda)_m}}}\right]^n}{n!} \quad (4.13)$$

$$= e^{-(P\lambda)_m} e^{(P\lambda)_m} e^{\frac{it_m}{\sqrt{2(P\lambda)_m}}}$$

$$E\{e^{\frac{it_m y_m}{\sqrt{2(P\lambda)_m}}}\} E\{e^{\frac{-it_m \vartheta_m}{\sqrt{2(P\lambda)_m}}}\} = e^{2(P\lambda)_m\left[\cos(\frac{t_m}{\sqrt{2(P\lambda)_m}})-1\right]} \quad (4.14)$$

$$= e^{2(P\lambda)_m\left[\sum_{n=1}^{+\infty}\frac{(-1)^n}{(2n)!}\left(\frac{t_m}{\sqrt{2(P\lambda)_m}}\right)^{2n}\right]}$$

$$\phi_{\xi}(t_1,\ldots,t_M) = e^{\left[2\sum_{m=1}^{M}(P\lambda)_m \sum_{n=1}^{+\infty}\frac{(-1)^n}{(2n)!}\left(\frac{t_m}{\sqrt{2(P\lambda)_m}}\right)^{2n}\right]} \tag{4.15}$$

By examining the first terms in the expansion of the power of the exponent,

$$\phi_{\xi}(t_1,\ldots,t_M) = e^{\frac{-1}{2}\sum_m t_m^2}\left[e^{O\left(\sum_k \frac{t_k^4}{48(P\lambda)_k}\right)}\right] \tag{4.16}$$

where $O(x)$ represents terms tending to zero at least as fast as x. As the mean number of counts in each term $(P\lambda)_j \rightarrow +\infty$, the joint characteristic function of ξ tends to that of M independent standard normal variables. Therefore, $\psi_s^2 = \sum_m \xi_m^2$ is asymptotically chi-squared with M degrees of freedom.

Chapter 5

GEM MAP Algorithms for the Poisson Data Model

5.1 Introduction

In image processing, it is held that images are well represented as consisting of segmented, relatively smooth regions [61]. This agrees with the concept of appropriate solutions in many medical image processing applications, such as emission tomography. We would like some basis for combining these qualitative insights with the data to arrive at an estimate of the undegraded image.

Two methods of incorporating this structural insight in image reconstruction are regularization and Bayesian estimation. Various forms of each of these two approaches have been suggested in the restoration and reconstruction literature. A regularization approach based on Grenanders method of sieves [30] was taken by Snyder and Miller in [71] and Miller, Snyder, and Moore in [54]. Bayesian reconstruction approaches have been examined by a number of authors. Recently, several authors have investigated Gaussian and Poisson priors in applying the EM approach to generate Bayesian reconstruction algorithms. Liang and Hart examined uncorrelated Gaussian and Poisson priors in [32] and in section 2 of [49] and correlated Gaussian priors in section 3 of [49]. The algorithms developed in those sections initially follow an EM derivation. However, in order to develop a closed form M-step, they replace a set of parameters in the prior distribution with a set of uncomputed image pixels values. Since this does not result in a true EM algorithm, as noted in [48], the question of whether the posterior distribution is increased at each step and the question of convergence for all possible data sets

remain open. Their algorithm was shown to out-perform the EM likelihood algorithm in a 1-dimensional simulation in [49] and later in 2-D simulations in [32]. Levitan and Herman [48] derive a valid EM algorithm for an uncorrelated Gaussian prior and demonstrate a marked improvement over ML estimates in 2-D simulations. For the mean of the prior distribution, Levitan and Herman used a smoothed filtered-backprojection reconstruction. Since it is generally accepted that the structure of images is one of non-stationary mean and local correlations [61], their prior satisfies the first of these two important image attributes.

Markov random fields capture the property of local correlation and offer a valuable image model. Non-iterative restoration of blurred images with additive i.i.d. Gaussian noise was addressed in [13] and [40] using the Gauss Markov image model of section 3.3.1. This model was coupled with a line process in [66] by Simchony, Chellappa, and Lichtenstein and the graduated non-convexity algorithm [7] was used to obtain globally optimal MAP estimates. In these papers the Markov random field model parameters are estimated from the noisy image. Cross and Jain [16] showed some of the textures that result from anisotropic Markov random field models, i.e. models of the form in section 3.3.1 for which the parameters θ_r are not equivalent. It is not clear to what degree an image restoration might benefit from the use of an anisotropic prior model for the image. In this chapter, general prior models are developed for images. The image is assumed to consist of varied content which would not in general contain a single global patterned variation such as that exhibited by an anisotropic Markov random field. The appropriate image model is taken to be an isotropic one.

A MAP approach based on a particular Gibbs prior, examined later in this chapter, was taken by Geman and McClure. In [23] they impose the constraint that

the unknown image pixels take values from a known interval and use gradient ascent to arrive at 2-D MAP estimates. Geman and Geman [22] use a coupled Markov random field state space combining intensity values and a line process which breaks the local dependence of a pixel on one or more of its neighbors. They employ simulated annealing to obtain MAP estimates. Gamble and Poggio [21] use a similar model and a Gibbs sampler to generate estimates which approximate the posterior mean.

In [5], Besag reported that his iterated conditional modes (ICM) algorithm, discussed in section 2.3, had been applied in a preliminary way to gamma camera scans which obey a Poisson data model. The ICM algorithm in section 2.3 updates a single pixel at a time by maximizing a univariate function which is conditioned on all data values to which that pixel contributes, all pixels which contribute to those data values, and all neighbors of the pixel to be updated. Where the image is degraded by a blurring function, each source pixel may contribute to many data values so that this algorithm can be more computationally demanding than the algorithm presented in this chapter.

In this chapter, we develop two GEM algorithms for MAP restoration and reconstruction of noisy blurred images based upon locally correlated Markov random field priors in the form of Gibbs functions. The first algorithm is specifically formulated for use only with the simplest Gibbs potential function and a 1^{st} or 2^{nd} order neighborhood. The second algorithm can be used with any Gibbs prior and any neighborhood. These algorithms are applied to the Poisson data model of section 2.2 . For the M-step of this GEM MAP algorithms, a form of coordinate gradient ascent is derived. Implementation of these algorithms closely follows that of the EM maximum likelihood (EM ML) algorithm [63]. In addition, as the prior

tends towards a uniform distribution, these algorithms reduce to the EM ML algorithm. The reconstructed image pixels are constrained to be non-negative and finite, but are not restricted to a specified finite interval as in [23] and no approximations as in [32],[49] are used. In this chapter, we apply these GEM MAP algorithms to estimate the 3-D image parameters in the Poisson model of single photon emission data. Three different Gibbs function priors are examined. Through these priors we examine one of the problems common to Bayesian image restoration based on Markov random field models. If the image model insures neighboring pixels take similar values, the restoration assumes an over-smoothed quality. If the model allows this bonding between neighboring pixels to break, either through a line process as in [21],[22],[66] or through the Gibbs potential function as in [23], some pixels assume values differing widely all their neighboring pixels. This becomes more apparent at lower SNRs at which these single pixels stand out in the restoration as either bright or dark points within regions of moderate intensity.

The generalized EM algorithm for MAP estimation was formulated in section 3.4.1.2. Optimization of likelihoods or Bayesian functions with independent priors may result in closed forms for the EM algorithm M-step [48]. If the complete data $\mathbf{x}$ are independent, the complete data are a linear function of the incomplete data, and the image pixels λ are treated as independent, the M-step only requires optimization of a set of univariate functions. However, it is generally accepted that the structure of images is one of local correlation [61]. It would therefore seem more desireable to examine the use of locally correlated priors. When used with an EM approach, correlated priors prohibit the existance of closed form solutions for an EM M-step. Each EM M-step thus requires an iterative optimization of an N-dimensional function, N being the dimension of λ. What results is an iterative optimization algorithm within each iteration of an iterative optimization algorithm.

In this case, iterative maximization of the posterior distribution by a method such as ICM [5] without using an EM formulation would initially seem more sensible than an EM approach. However, the GEM algorithms presented in this paper offer an attractive alternative to both of these approaches. A true EM formulation would require a global optimization in the M-step of the algorithm, which can be very computationally demanding to implement. We prefer a generalized EM approach since it guarantees a monotonic increase of the posterior function and has proven convergence properties. Under this approach, each M-step may consist of 1 or more iterations of an algorithm to increase $Q(\lambda|\lambda^k)$, the result from the E-step of the generalized EM algorithm, without the requirement of maximizing it. We would expect the per-iteration speed of convergence to be slower for a GEM versus an EM approach. However, where a closed form M-step does not exist, the GEM approach may generate estimates with greater increases in the posterior distribution function for a given amount of computation. This may often be the case since it is generally true that the first iteration of an iterative optimization algorithm produces the largest improvement.

It should be noted that the use of a non-uniform prior distribution can induce local minima as well as local maxima in both the posterior distribution and in the results from an E-step of an EM formulation. Globally optimal solutions can be obtained by applying the simulated annealing algorithm (section 3.4.2.1); however the cooling schedule must be very slow, requiring an extremely large number of iterations. Deterministic algorithms, such as those presented in this chapter, at best converge to locally optimal estimates. However, the benefit derived by these algorithms is a reduction in computational requirement per iteration with many fewer iterations being required to achieve convergence. This becomes increasingly important as the size of the restoration or reconstruction problem increases or when

results are desired as soon as possible after the data images have been stored. We apply these algorithms both to image restoration, in which the data image has been degraded both by blur and by noise, and to 3-D image reconstruction for the emission tomography model.

5.2 Gibbs Prior Potential Functions

The algorithms presented in the following sections are formulated for MAP estimation based on a Gibbs distribution prior. A Gibbs distribution is a probability measure on the set of configurations $\{\lambda\}$ which has the form

$$f(\lambda) = \frac{1}{K_\beta} e^{\frac{-U(\lambda)}{\beta}} \tag{5.1}$$

where β is the Gibbs prior parameter, K_β is the normalizing constant (partition function), and $U(\lambda)$ is termed the energy function. The energy function has the form

$$U(\lambda) = \sum_{c \in C} \sum_{I_c} V_c(\lambda_i .. \lambda_m : \{i..m\} \in I_c) \tag{5.2}$$

where C denotes the set of all cliques types, I_c a set of lattice sites in a clique of type c, and $V_c(\lambda)$, termed a potential function, is a function of the values at the sites in I_c. To form the Gibbs prior we select a neighborhood structure and define the potential functions defined on the cliques for the chosen neighborhood. In applying the algorithms which are derived in the following sections, we have chosen a 2nd order neighborhood for application to 2-D images and a 1st order neighborhood for application to 3-D images (section 3.3). In [5] a 1st order neighborhood is considered unrealistic for most applications. However, in three dimensions a 1st order neighborhood results in 6 neighbors for every pixel versus 4 in

two dimensions. As results presented later indicate, a 1st order neighborhood in three dimension may be sufficient for many applications. The potential function for cliques containing only a single pixel has been set equal to zero. In addition, for 2nd order neighborhoods, cliques containing three pixels have been set equal to zero. For all cliques containing two pixels, the same potential function is used; but for cliques containing diagonally offset pixels the potential functions are normalized by $\sqrt{2}$. Therefore, we only use a single potential function to define Gibbs distribution. The three potential functions we examine are:

$$(1)\quad V_1(\lambda_j;\lambda_i) = (\lambda_j - \lambda_i)^2 \tag{5.3}$$

$$(2)\quad V_2(\lambda_j;\lambda_i) = \frac{(\lambda_j-\lambda_i)^2}{\delta^2 + (\lambda_j - \lambda_i)^2} \tag{5.4}$$

$$(3)\quad V_3(\lambda_i;\lambda_j) = \log\left[1+\left(\frac{\lambda_j-\lambda_i}{\mu}\right)^2\right] \tag{5.5}$$

The first, equation (5.3), is used by Geman and Geman in [22]. The second, equation (5.4), is equivalent to that used by Geman and McClure in [23]. The third potential function, equation (5.5), is a compromise between the first two. A plot of these three potential functions, suitably normalized, versus the difference between the two pixels in a clique is shown in figure 5.1.

This normalization of the three potential functions simplifies the interpetation of the effects of the priors and enables some comparison for a given β value. The first potential function increasingly penalizes the separation between neighboring pixels. In addition, it does so at an increasing rate as the separation increases. To improve on this, we seek a potential function which penalizes separations within uniform regions without unduly penalizing the larger separations which we foresee occuring at the boundary between two different regions of the image. The second

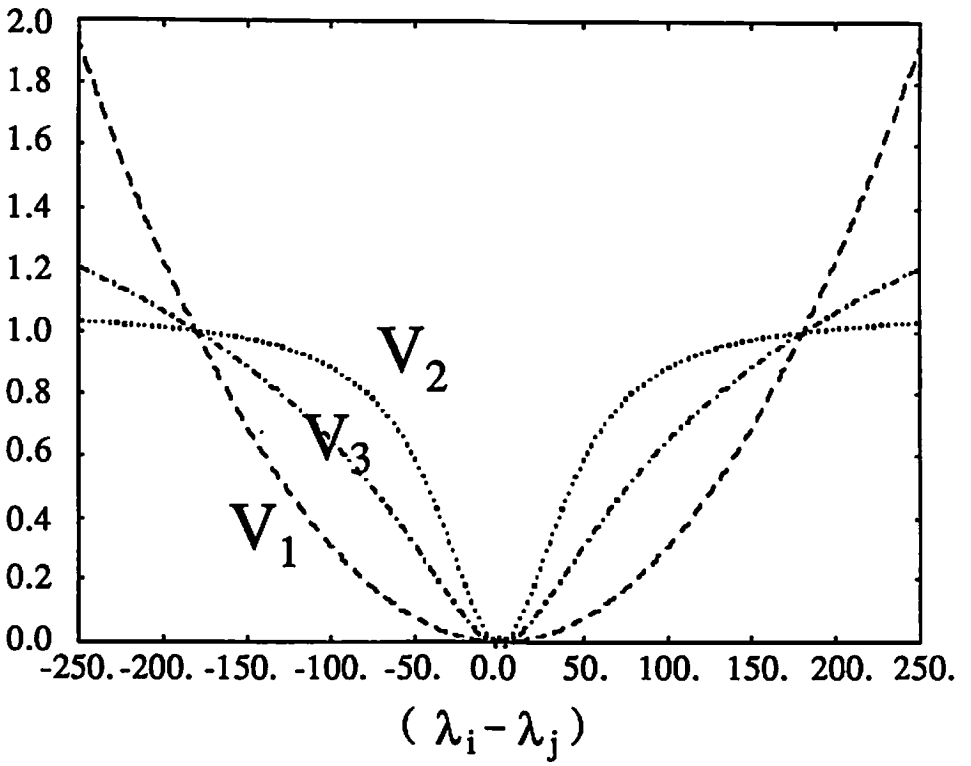

Figure 5.1
Plot of the three potential functions versus separation of neighboring pixels.

prior encourages neighboring pixels to be of similar value until they have become separated through iterative reconstruction by a given threshold δ. At this point the prior allows furthur separation of their values at relatively small increase in the penalty. We found that in MAP reconstructions of single photon emission images, some single pixels may separate in value from all their neighboring pixels. When this occurs, the second prior allows this separation to be exaggerated. We found that the quality of the reconstruction could benefit from a compromise between the first and second priors. The third prior represents a compromise which alleviates this problem by increasingly penalizing pixel separation.

5.3 The Generalized EM Approach with Poisson Data

The Poisson data model (section 2.2) assumes the data y to be conditionally independent and Poisson distributed with mean equal to a blurred version $\mathbf{P}\boldsymbol{\lambda}$ of the uncorrupted image $\boldsymbol{\lambda}$. The uncorrupted image $\boldsymbol{\lambda}$ is allowed to take the complete range of non-negative values $\{\boldsymbol{\lambda} \geq \boldsymbol{\lambda}\}$. Let $\mathbf{P}\boldsymbol{\lambda}$ be the blurred version of the uncorrupted image $\boldsymbol{\lambda}$. The data y are Poisson so that their distribution conditioned upon the uncorrupted image $\boldsymbol{\lambda}$ is

$$f(\mathbf{y}\,|\,\boldsymbol{\lambda}) = \prod_i e^{-(\sum_j \mathbf{P}_{ij}\boldsymbol{\lambda}_j)} \frac{(\sum_j \mathbf{P}_{ij}\boldsymbol{\lambda}_j)^{\mathbf{y}_i}}{(\mathbf{y}_i)!} . \tag{5.6}$$

It was pointed out by Shepp and Vardi [64] that this problem has a more natural formulation in terms of a set of complete, but unobserved data. This complete/incomplete data formulation allows the use of the expectation - maximization approach discussed in section 3.4.1.2. Shepp and Vardi applied this formulation to derive a maximum likelihood algorithm for emission tomography which has enjoyed widespread popularity for solution of this reconstruction problem. Consider a set of intermediate Poisson data variables $\{\mathbf{x}_{ij}\}$ which are not directly measureable. In the terminology of Dempster, Laird, and Rubin [18] these intermediate variables will be the complete data set. Let the intermediate data value $\mathbf{x}_{ij}$ represent the random contribution of the j^{th} uncorrupted image pixel to the i^{th} data value. It therefore has mean mean $\mathbf{P}_{ij}\boldsymbol{\lambda}_j$. Clearly, $\mathbf{y}_i = \sum_i \mathbf{x}_{ij}$. and $E\{\mathbf{y}_i\} = \sum_j \mathbf{P}_{ij} E\{\mathbf{x}_{ij}\}$. The complete data x are conditionally independent so that

$$f(\mathbf{x}|\lambda) = \prod_i \prod_j e^{-(P_{ij}\lambda_j)} \frac{(P_{ij}\lambda_j)^{x_{ij}}}{(x_{ij})!} . \tag{5.7}$$

The logarithm of $f(\mathbf{x}|\lambda)$ is a linear function of the x_{ij}'s plus terms independent of λ. Therefore, to compute the E-step, equation (3.26), we only need to compute terms containing $E\{x_{ij}|y\lambda^k\}$. The variables $\{x_{iq}:q = 1,..,N\}$ are conditionally independent and Poisson with means $P_{iq}\lambda_q$. In addition, $\sum_q x_{iq} = y_i$ so that $\{x_{i1}..x_{iN}\}$ are only dependent upon data value y_i, whereby

$$f(x_{i1}..x_{iN}|y;\lambda^k) = f(x_{i1}..x_{iN}|y_i;\lambda^k) \tag{5.8}$$

The joint distribution of a set of independent Poisson variables z conditioned on their sum is a multinomial distribution with probability in each class j equal to $\frac{E\{z_j\}}{\sum E\{z_j\}}$ [58]. The expected value of a multinomial variable which represents the number of occurences of its class is equal to the probability in its class times the total number of trials. The joint distribution of $\{x_{iq}:q=1,..N\}$ conditioned on $\sum_q x_{iq}=y_i$ can therefore be formed as

$$f(x_{i1}..x_{iN}|y_i;\lambda^k) = \frac{(y_i)!}{(x_{i1})!\cdots(x_{iN})!}(\mu_{i1})^{x_{i1}}\cdots(\mu_{iN})^{x_{iN}} \tag{5.9}$$

$$\text{where} \quad \mu_{ij} = \frac{P_{ij}\lambda_j}{\sum_q P_{iq}\lambda_q} .$$

Therefore, $E\{x_{ij}|y\lambda^k\}$, which is the mean of class j of the multinomial distribution, is equal to $y_i\,\mu_{ij}$ which is equivalently $\frac{y_i P_{ij}\lambda_j^k}{\sum_q P_{iq}\lambda_q^k}$. The E-step can now be formed as

$$E_{\mathbf{x}}\{\log f(\mathbf{x}\,|\,\lambda)\,|\,\mathbf{y}\lambda^k\} = \sum_j(-\mathbf{a}_j\lambda_j + \mathbf{b}_j^k\log\lambda_j) + \text{terms independent of } \lambda$$

$$\text{where} \quad \mathbf{a}_j = \sum_i \mathbf{P}_{ij} \qquad \text{and} \qquad \mathbf{b}_j^k = \sum_i \frac{\mathbf{y}_i \mathbf{P}_{ij}\lambda_j^k}{\sum_q \mathbf{P}_{iq}\lambda_q^k} \tag{5.10}$$

If the $\mathbf{P}_{ij}$'s have been normalized as in [63],[64],[65] then $\mathbf{a}_j = 1$ for all j. In forming the E-step $Q(\lambda\,|\,\lambda^k)$ of equation (5.10) we can omit the terms independent of λ since these do not affect the M-step equation (3.27). For the EM approach, the M-step requires maximization of $Q(\lambda\,|\,\lambda^k)$ where, from equations (3.27) and (5.10),

$$Q(\lambda\,|\,\lambda^k) = \sum_j(-\mathbf{a}_j\lambda_j + \mathbf{b}_j^k\log\lambda_j) - \sum_{c\in C}\sum_{I_c}\frac{V_c(\lambda_i..\lambda_m:\{i..m\}\in I_c)}{\beta} \tag{5.11}$$

However, a closed form EM M-step does not exist in this case so that a single EM M-step would require an iterative solution to the maximization problem posed by the M-step. Instead, we propose a generalized EM approach [18],[84] employing one or more cycles of coordinate ascent [52] for the M-step.

5.4 A GEM Algorithm Specific to V_1

For the generalized EM approach, the M-step is to find a λ^{k+1} such that $Q(\lambda^{k+1}\,|\,\lambda^k) > Q(\lambda^k\,|\,\lambda^k)$. Coordinate ascent is an optimization approach which maximizes a functional by changing a single component of the argument of the functional at a time. Each component is addressed in some prescribed order so that for bounded functions ultimate convergence to a stationary value is assured.

The algorithm presented in this section is formulated specifically for the potential function V_1 in equation (5.3). For the M-step, each pixel can be updated in turn such that each new pixel value produces the largest possible increase in $Q(\lambda\,|\,\lambda^k)$.

Let N_j represents the set of indices of the pixels which are neighbors of pixel j. A pixel λ_j contributes to $Q(\lambda|\lambda^k)$, equation (5.11), only through the terms

$$q_j(\lambda_j) = -a_j\lambda_j + b_j^k \log\lambda_j - \frac{1}{\beta}\sum_{i\varepsilon N_j}(\lambda_j - \lambda_i)^2. \tag{5.12}$$

Setting the derivative of $q_j(\lambda_j)$ equal to zero to find the two extrema results in a second order polynomial in λ_j.

$$\lambda_j^2\left(\frac{1}{\beta}\sum_{i\varepsilon N_j} 2\right) + \lambda_j\left(a_j - \frac{1}{\beta}\sum_{i\varepsilon N_j}\lambda_i\right) - b_j^k = 0 \tag{5.13}$$

Let m_j equal the number of neighbors of pixel j and $c = \frac{\beta a_j}{2m_j} - \frac{1}{m_j}\sum_{i\varepsilon N_j}\lambda_i$. Then the two roots r_i are

$$r_i = \frac{1}{2}\left[-c \pm \sqrt{c^2 - 2\beta b_j^k/m_j}\,\right] \tag{5.14}$$

Since all elements of vector λ^k are non-negative, $b_j^k \geq 0$. In addition, β and m_j are greater than 0. Therefore the root formed by subtracting the radical term is always < 0 unless both roots are identically zero. This only occurs if b_j^k is zero. The second derivative of $q_j(\lambda_j)$ is

$$\frac{\partial^2 q_j(\lambda_j)}{\partial\lambda_j^2} = -\left(\frac{b_j^k}{\lambda_j^2} + \frac{2}{\beta}\sum_{i\varepsilon N_j}\lambda_j\right) \tag{5.15}$$

Over the range $\lambda_j \geq 0$, b_j^k is greater than 0, so that the second derivative of $q_j(\lambda_j)$ is less than zero there. If the root formed by adding the radical term in equation (5.14) is ≥ 0 then $q_j(\lambda_j)$ is maximized for λ_j^{k+1} equal to that value. If that root is negative, the the derivative of $q_j(\lambda_j)$ is negative for all $\lambda_j^{k+1} \geq 0$ and $q_j(\lambda_j)$ is maximized over the non-negative values by λ_j^{k+1} equal to zero.

The image pixels are updated in any sequential order with one or more updating cycle defining the M-step. We have found, that updating each image pixel only twice in the M-step generally produces the greatest convergence for a given computation time.

The simulation we used to test this algorithm is a model of a 3-D single photon emission imaging system consisting of a parallel collimated gamma camera with 48^2 pixels viewing a 3-D volume with data collected from 48 different equispaced angles. Perfect collimation is assumed. A 3-D source space consisting of 48^3 pixels is reconstructed. A 3-D computer generated head phantom shown in figure 5.2 was used as the uncorrupted image.

The Gibbs distribution potential function V_1 was defined as in equation (5.3). A 1^{st} order neighborhood in 3-D consisting of the 6 nearest neighbors of an image pixel was used. The neighbors of an interior image pixel consist of the pixels above, below, and on all 4 sides totaling 6 neighbors. The missing neighbors of pixels located on the side boundaries are assumed zero since this is the case in emission tomography. This is due to the physical geometry of the imaging system. A free boundary [22] is used for pixels on the top and bottom planes of the 3-D reconstruction space. These pixels have fewer neighbors in that pixels on the bottom layer of the 3-D reconstruction space have no neighbor below while pixels on the top layer have no neighbor above.

Figure 5.3 shows the L_2 error performance of this GEM MAP algorithm versus that of the EM ML algorithm for 100 iterations of each algorithm. This plot demonstrates the potential improvement in reconstructions using a MAP approach with the Gibbs prior V_1 of equation (5.3). In addition, it shows that the GEM MAP algorithm converged in fewer iterations since the L_2 error from the EM ML

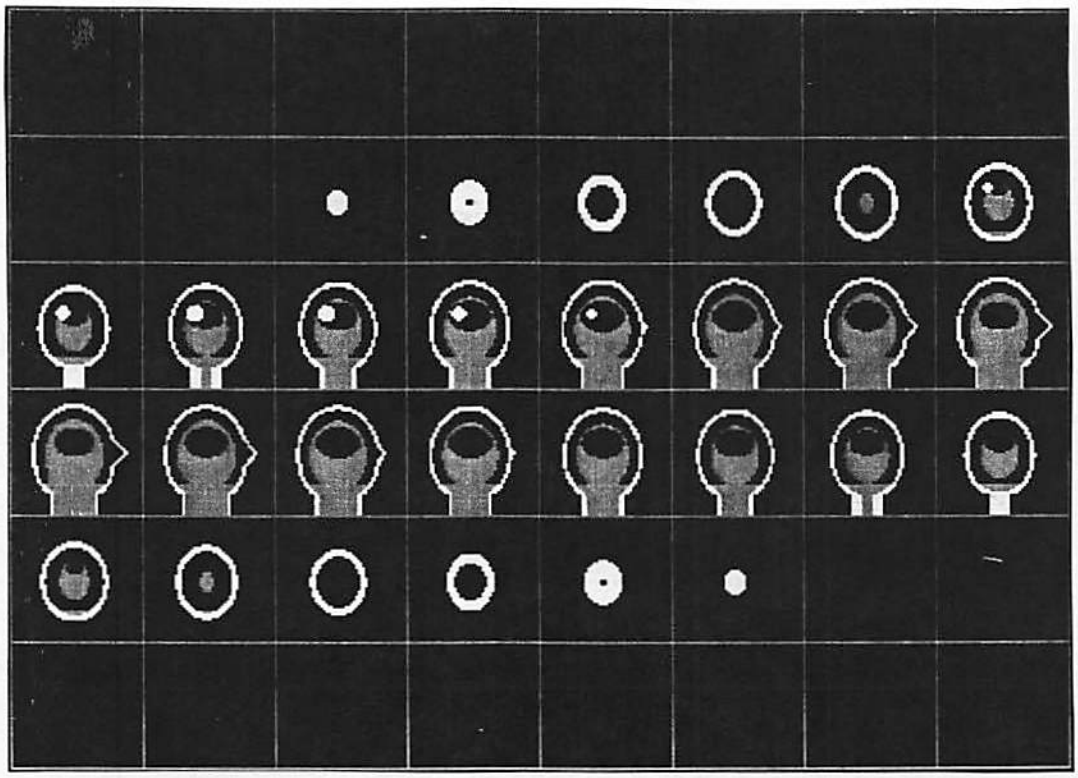

Figure 5.2
The computer generated 3-D head phantom consisting of 48 longitudal planes. The upper left-hand image shows the plane nearest the viewer, with subsequent planes shown across the top row, then row by row.

algorithm is still increasing after 100 iterations. Figure 5.4 demonstrates that the Bayes function $B(\boldsymbol{\lambda} \mid \mathbf{y})$ is increased at each iteration of this GEM MAP algorithm.

5.5 A GEM Algorithm for any Gibbs Prior

The next algorithm is more general in that it can be applied with any Gibbs prior. It is a generalized expectation-maximization (GEM) algorithm for maximum a posteriori image restoration and reconstruction. This algorithm is therefore referred to as a GEM MAP algorithm. Its implementation closely follows that of

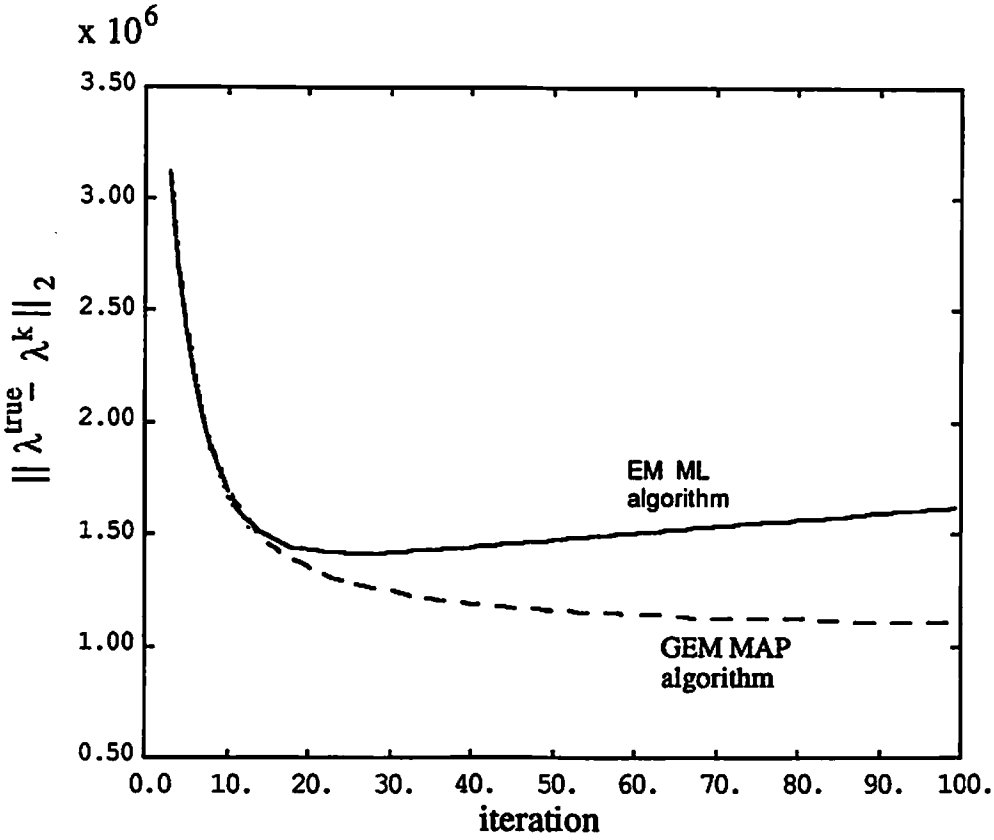

Figure 5.3
Plot of the L_2 difference between the true 3-D image and the reconstruction versus 100 iterations of the EM ML algorithm and the first GEM MAP algorithm for potential function V_1.

the EM ML algorithm [64]. We note that this algorithm updates pixels sequentially and that updated values are used to update the pixels that follow. Therefore, in step (2a) below, the superscript has been omitted from λ_i in $V(\lambda_i;\lambda_j^k)$ because the neighboring pixels λ_i may consist of both updated and un-updated pixels. Step (1) contains the E-step of this GEM MAP algorithm and step (2) contains the M-step. Step (2) can be repeated so that each pixel is visited and updated more than once during a GEM MAP iteration. Generally, step (2) need only be performed once or

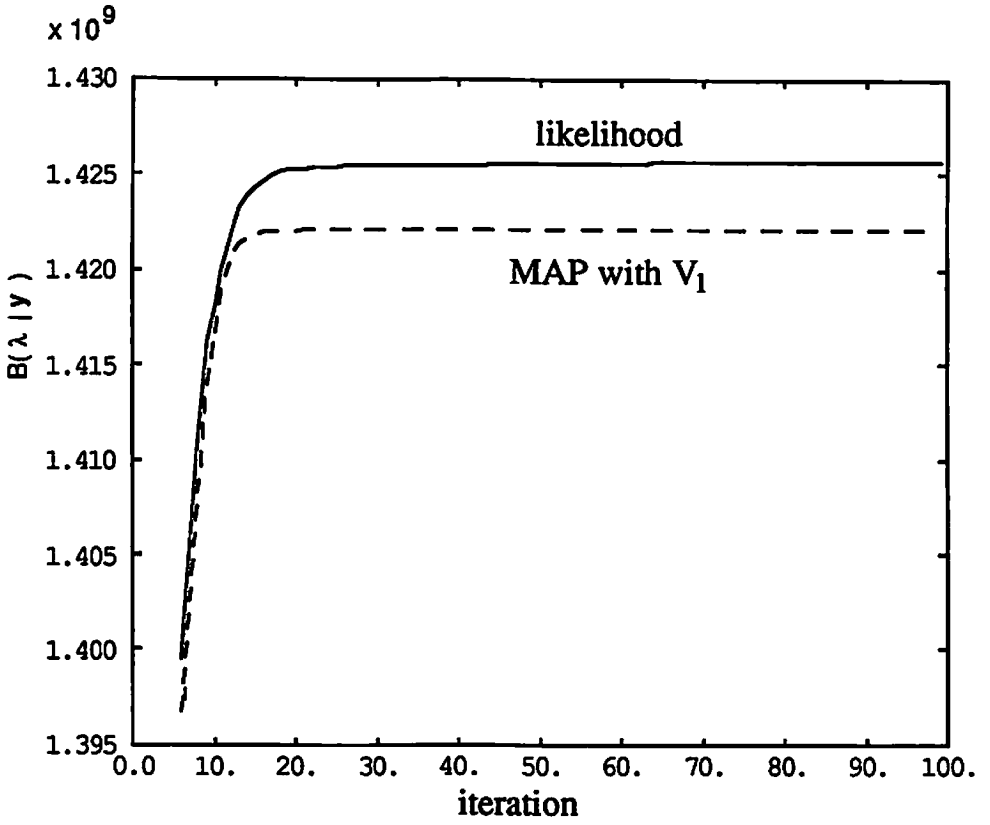

Figure 5.4
Plot of the likelihood and Bayes function $B(\lambda | y)$ versus 100 iterations of the EM ML and the GEM MAP algorithms with V_1.

twice after each completed step (1) to achieve an expeditious rate of convergence. At the k^{th} iteration, perform the following steps:

(1) For all image pixels, compute the usual EM maximum likelihood algorithm [61] updated variables $\lambda_j^{EM} = \frac{b_j^k}{a_j}$ where a_j and b_j^k are defined in equation (5.10).

(2) To update the image, visit pixel sites sequentially. When visiting a pixel λ_j, do (2a)-(2d)

(2a) compute C_1 and C_2 where

$$C_1 = \mathrm{a}_j\left(-\lambda_j^k + \lambda_j^{EM}\log\lambda_j^k\right) - \sum_{c\varepsilon C}\sum_{I_c : j\varepsilon I_c}\frac{V(\lambda_j^k;\lambda_i..\lambda_m:\{ji..m\}\varepsilon I_c)}{\beta}$$

$$C_2 = \frac{1}{\beta}\sum_{c\varepsilon C}\sum_{I_c : j\varepsilon I_c}\frac{\partial}{\partial\lambda_j^k}V(\lambda_j^k;\lambda_i..\lambda_m:\{ji..m\}\varepsilon I_c)$$

(2b) Set $\alpha = 1$.

Compute $\lambda_j^{k+1} = \lambda_j^{\mathrm{EM}} - \dfrac{C_2\lambda_j^k}{\mathrm{a}_j}$

If $\lambda_j^{k+1} > 0$ go to (2d)

If $\lambda_j^{k+1} \le 0$ compute $\alpha = \dfrac{.5}{1 - \dfrac{\lambda_j^{\mathrm{EM}}}{\lambda_j^k} + \dfrac{C_2}{\mathrm{a}_j}}$.

(2c) Compute $\lambda_j^{k+1} = (1-\alpha)\lambda_j^k + \alpha\left\{\lambda_j^{EM} - \dfrac{C_2\lambda_j^k}{\mathrm{a}_j}\right\}$

(2d) Check if

$$\mathrm{a}_j\left(-\lambda_j^{k+1} + \lambda_j^{EM}\log\lambda_j^{k+1}\right) - \sum_{c\varepsilon C}\sum_{I_c : j\varepsilon I_c}\frac{V(\lambda_j^{k+1};\lambda_i..\lambda_m:\{ji..m\}\varepsilon I_c)}{\beta} \ge C_1$$

If yes, update pixel j to λ_j^{k+1} visit the next pixel. If no, divide α by 2 and return to step (2c).

To further clarify the steps, when visiting a pixel j, step (2a) computes two values which are functions of the data and of the updated and un-updated pixels in the neighborhood of pixel j. In step (2b), if the step-size $\alpha = 1$ results in a

negative λ_j^{k+1}, half the step-size α which gives $\lambda_j^{k+1} = 0$ is computed. Step (2c) implements a coordinate ascent [48] step for the specified step-size, and step (2d) ensures that the step-size α has resulted in an increase in $Q(\lambda|\lambda^k)$. If $Q(\lambda|\lambda^k)$ has not been increased, step (2d) cuts the step-size in half. As $\beta \rightarrow +\infty$, $C_2 \rightarrow 0$ and the pixels are updated by setting them equal to the EM ML updated pixel values. Step (2d) , which would then represent a check to ensure the likelihood function has been increased, is always satisfied so that steps (2a), (2c), and (2d) are no longer necessary. As $\beta \rightarrow +\infty$, this algorithm thus reduces to the EM ML algorithm.

Let us consider why this is a GEM algorithm. If step (2d) ensures a monotonic increase of $Q(\lambda|\lambda^k)$ and step (2c) is guaranteed to arrive at a λ_j^{k+1} satisfying step (2d), then at the conclusion of the M-step $Q(\lambda^{k+1}|\lambda^k) > Q(\lambda^k|\lambda^k)$ and the algorithm is a GEM algorithm. Without loss of generality, consider the case of a 1st order neighborhood for which the potential function on cliques containing a single pixel are set to zero. Specify a potential function $V(\lambda_i;\lambda_j)$ evaluated on all cliques containing two pixels. The energy function has the form $\frac{1}{\beta}\sum_{ij \in \mathcal{L}_{c_2}} V(\lambda_i;\lambda_j)$. Let N_j denote the set of indices of pixels which are neighbors of pixel j. In order to show that equation (5.11) is increased at each stage, there are only two cases we must consider: updating a pixel whose neighbors have not been updated in the present iteration, and updating a pixel for which one or more of the neighbors have been updated. Without loss of generality, let us examine steps (2a)-(2d) for two pixels λ_i and λ_j which are neighbors by writing $Q(\lambda|\lambda^k)$ explicitly in terms of λ_i and λ_j.

$$Q(\lambda_i;\lambda_j;\lambda_l: l\neq i,j \mid \lambda^k) = -a_i\lambda_i + b_i^k \log\lambda_i - \sum_{\substack{l\in N_i \\ l\neq j}} \frac{V(\lambda_i;\lambda_l)}{\beta} - \frac{V(\lambda_i;\lambda_j)}{\beta} \qquad (5.16)$$

$$-a_j\lambda_j + b_j^k \log\lambda_j - \sum_{\substack{l\in N_j \\ l\neq i}} \frac{V(\lambda_j;\lambda_l)}{\beta} + \sum_{l\neq i,j} (-a_l\lambda_l + b_l^k \log\lambda_l) - \sum_{\substack{pq\in C \\ p;q\neq i \text{ or } j}} \frac{V(\lambda_p;\lambda_q)}{\beta}$$

Let us first visit pixel i whose neighbors have not been updated. Step (2d) ensures an increase in the sum of terms 1, 2, 3, and 4 in equation (5.16), with no effect on the other terms. Therefore, λ_i^{k+1} satisfies

$$Q(\lambda_i^{k+1};\lambda_j^k;\lambda_l^k: l\neq i,j \mid \lambda^k) > Q(\lambda_i^k;\lambda_j^k;\lambda_l^k: l\neq i,j \mid \lambda^k) \qquad (5.17)$$

For the second case, let us then visit and update pixel j a neighbor of pixel i , step (2d) furthur guarantees an increase in the sum of terms 4, 5, 6, and 7 in equation (5.16), with no effect on the other terms. Therefore, λ_j^{k+1} satisfies

$$Q(\lambda_i^{k+1};\lambda_j^{k+1};\lambda_l^k: l\neq i,j \mid \lambda^k) > Q(\lambda_i^{k+1};\lambda_j^k;\lambda_l^k: l\neq i,j \mid \lambda^k) \qquad (5.18)$$

Each pixel that is updated according to step (2d) results in an increase in $Q(\lambda \mid \lambda^k)$ and from section 3.4.1.2, $B(\lambda \mid y)$ is increased.

It remains to show that step (2c) will result in an updated pixel satisfying step (2d). Step (2c) is a coordinate gradient ascent [52] of $Q(\lambda \mid \lambda^k)$ initialized at λ^k . At each step of a coordinate ascent algorithm, only changes to a single element λ_j are allowed such that a monotonic increase in $Q(\lambda \mid \lambda^k)$ is achieved. Each element is addressed in some prescribed order so that for bounded functions ultimate convergence to a stationary value is assured. Let us take the case above, where λ_i^k has been updated to λ_i^{k+1} and we wish to update λ_j^k. A coordinate gradient ascent of $Q(\lambda_i^{k+1};\lambda_j^k;\lambda_l^k: l\neq i,j \mid \lambda^k)$ takes the form

$$\lambda_j^{k+1} = \lambda_j^k + \mathbf{m}_j \frac{\partial}{\partial \lambda_j^k} Q(\lambda_i^{k+1}; \lambda_j^k; \lambda_l^k : l \neq i, j \mid \lambda^k) \tag{5.19}$$

where $\mathbf{m}_j$ is any positive value. From equation (5.15)

$$\frac{\partial}{\partial \lambda_j^k} Q(\lambda_i^{k+1}; \lambda_j^k; \lambda_l^k : l \neq i, j \mid \lambda^k) = -\mathbf{a}_j + \frac{\mathbf{b}_j^k}{\lambda_j^k} - C_2 \tag{5.20}$$

$$= \mathbf{a}_j \left(-1 + \frac{\lambda_j^{EM}}{\lambda_j^k} - \frac{C_2}{\mathbf{a}_j}\right)$$

where C_2 is as defined in step (2a). Let $\mathbf{m}_j = \dfrac{\alpha \lambda_j^k}{\mathbf{a}_j}$. Since λ^k is constrained to be positive, $0 < \alpha \leq 1$, and $\mathbf{a}_j > 0$, $\mathbf{m}_j$ is positive. Substituting $\mathbf{m}_j$ and equation (5.20) into equation (5.19) gives step (2c). The step direction follows the j^{th} coordinate directional derivative of $Q(\lambda_i^{k+1}; \lambda_j^k; \lambda_l^k : l \neq i, j \mid \lambda^k)$ and the initial step-size is chosen to mimic the EM likelihood algorithm. An increase in $Q(\lambda_i^{k+1}; \lambda_j^k; \lambda_l^k : l \neq i, j \mid \lambda^k)$ is possible with a sufficiently small step size if that derivative is non-zero. If the derivative is zero, step (2d) is satisfied immediately. Steps sizes which do not satisfy step (2d) are halved so that the algorithm quickly arrives at a step size increasing $Q(\lambda \mid \lambda^k)$.

This GEM algorithm monotonically increases $Q(\lambda \mid \lambda^k)$ and terminates at a point λ^k for which

$$[\nabla Q(\lambda^k \mid \lambda^k)]_j \rightarrow \begin{Bmatrix} = 0 \text{ if } \lambda_j^k > 0 \\ < 0 \text{ if } \lambda_j^k = 0 \end{Bmatrix} \tag{5.21}$$

for all j; i.e. at a point where the directional derivative in all feasible directions is less than or equal to zero. From equation (3.23)

$$\nabla B(\lambda \mid y) = \nabla Q(\lambda \mid \lambda^k) - \nabla E_x\{\log f(x \mid y\lambda) \mid y\lambda^k\} \tag{5.22}$$

Since λ^k maximizes $E_x\{\log f(x \mid y\lambda) \mid y\lambda^k\}$, Jensens inequality section 3.4.1.2, $\nabla E_x\{\log f(x \mid y\lambda^k) \mid y\lambda^k\} = 0$ [84]. This holds regardless of whether or not λ^k lies on a boundary. Therefore $\nabla B(\lambda^k \mid y) = \nabla Q(\lambda^k \mid \lambda^k)$ and the algorithm terminates at a point λ^k for which the directional derivative of $B(\lambda^k \mid y)$ in any feasible direction is less than or equal to zero, i.e. $\{\lambda^k\}$ converges to a stationary point.

The simulation we used to test this algorithm is a model of a 3-D single photon emission imaging system consisting of a parallel collimated gamma camera with 48^2 pixels viewing a 3-D volume with data collected from 48 different equispaced angles. Perfect collimation is assumed. A 3-D source space consisting of 48^3 pixels is reconstructed. The 3-D computer generated head phantom shown in figure 5.2 was used as the uncorrupted image. The source space, consisting of 48^3 pixels, was reconstructed.

The Gibbs distribution potential functions were defined as in section 3 for a 1^{st} order neighborhood consisting of the 6 nearest neighbors. The neighbors of an interior image pixel consist of the pixels above, below, and on all 4 sides totaling 6 neighbors. The missing neighbors of pixels located on the side boundaries are assumed zero. A free boundary [22] is used for pixels on the top and bottom planes of the 3-D reconstruction space. These pixels have fewer neighbors. Pixels on the bottom layer of the 3-D reconstruction space have no neighbor below while pixels on the top layer have no neighbor above.

The reconstructions displayed in the following figures are the result of 50 iterations of each specified algorithm. With more iterations, the EM ML reconstruction deteriorated furthur while the GEM MAP reconstructions remained unchanged. For these reconstructions β=1 and the three potential functions were normalized as

shown in figure 5.1. Figure 5.5 shows the ML reconstruction using the EM ML algorithm of Shepp and Vardi [64]. Evident is the the excessive non-smoothness reported by many authors [23],[33],[34],[48],[63],[71]. By introducing prior information concerning the smoothness of the true underlying image, the reconstruction will exhibit the image qualities which are held representative of a distribution of tagged pharmaceuticals in a patient. Figures 5.6, 5.7, and 5.8 show 3-D reconstructions using the GEM MAP algorithm with the three Gibbs priors in equations (5.3), (5.4), and (5.5). Upon visual examination and comparison with the undegraded 3-D image of figure 5.2 and the maximum likelihood reconstruction of figure 5.5, all three Gibbs priors produced a considerable visible improvement in the reconstruction. Close examination of the MAP reconstruction of figure 5.6 which is based on a Gibbs prior with potential function V_1 in equation (5.3) shows that the edges in the image are more blurred than those in the maximum likelihood reconstruction. This is because the potential function heavily penalizes the disparity between neighboring pixels which occur at an edge. We would expect the MAP reconstruction in figure 5.7 which uses potential function V_2 in equation (5.4) to have much more distinct and well defined edges. As discussed previously this is because the potential function allows large disparities to occur between neighbors without a significant increase in the penalty as compared to that for much smaller disparities. However, in this reconstruction, figure 5.7, a number of single pixels can be seen to have taken very large values, although all their neighbors have much lower values. These single pixels stand out as points of high intensity in regions of moderate intensity. Alternatively, this same effect can result in single pixels of low intensity within regions of moderate intensity. This is the drawback of V_2 as previously mentioned. The MAP reconstruction in figure 5.8 using V_3 corrects this problem to the degree that it produces solutions slightly less smooth. This may be preferable in

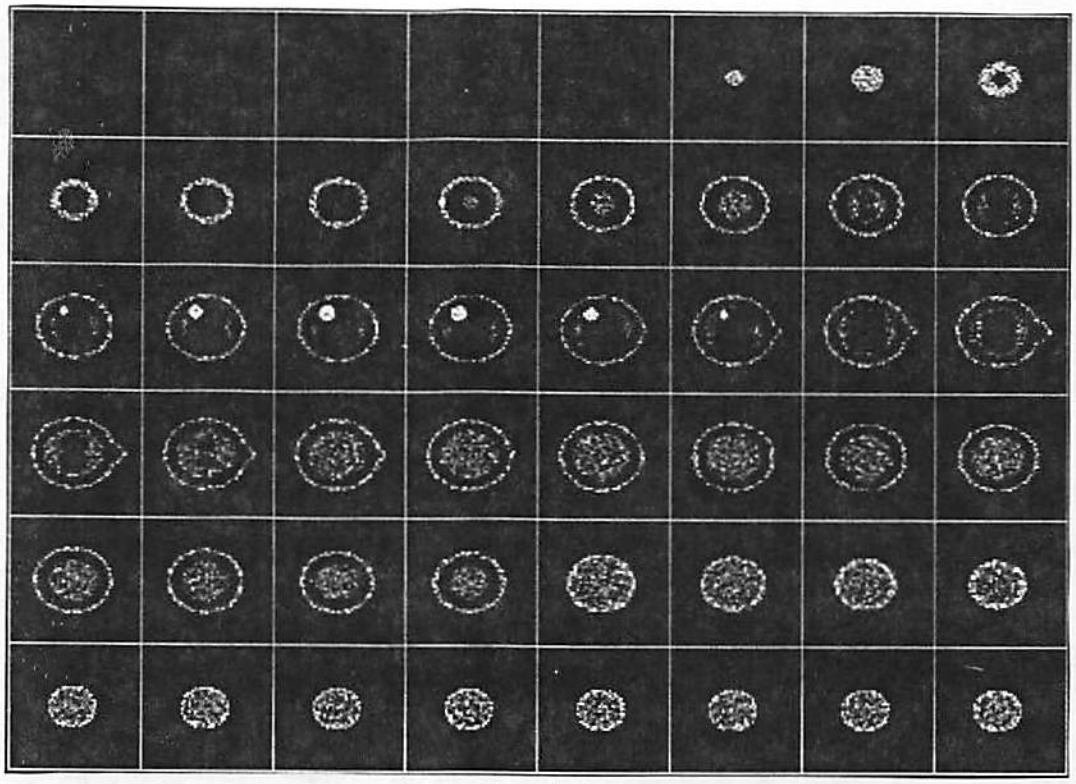

Figure 5.5
The ML reconstruction of the 3-D head phantom in figure 5.2.

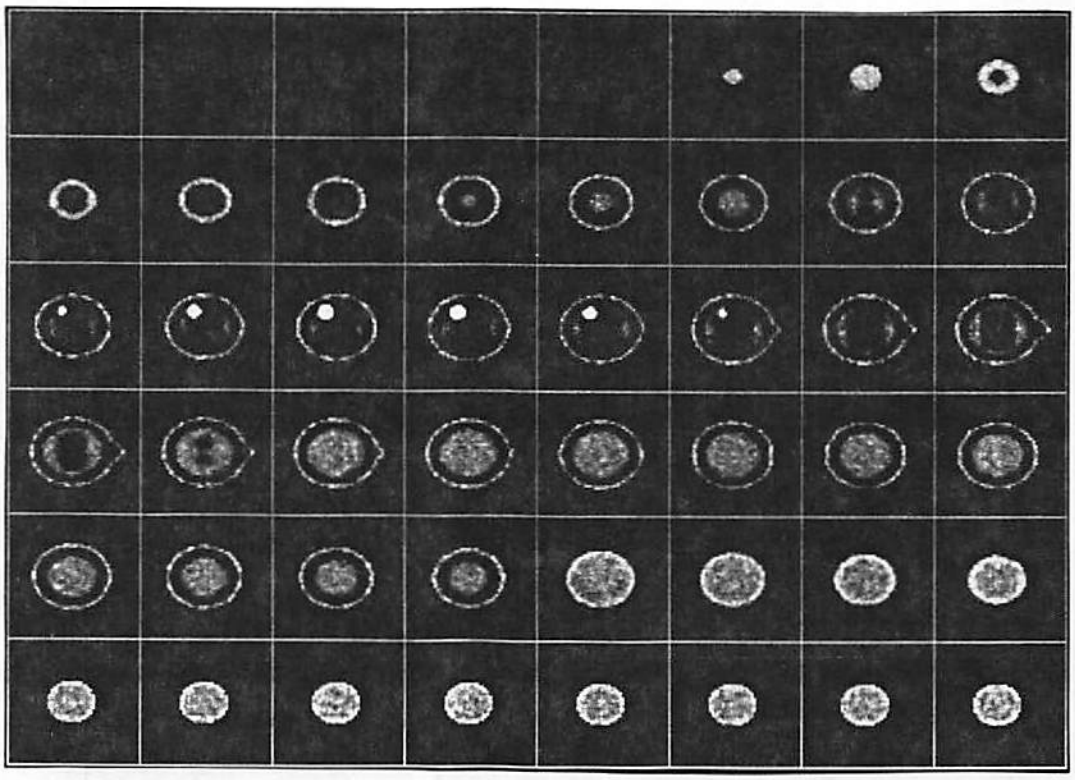

Figure 5.6
MAP reconstruction of the 3-D head phantom in figure 5.2 using Gibbs prior V_1 in equation (5.3).

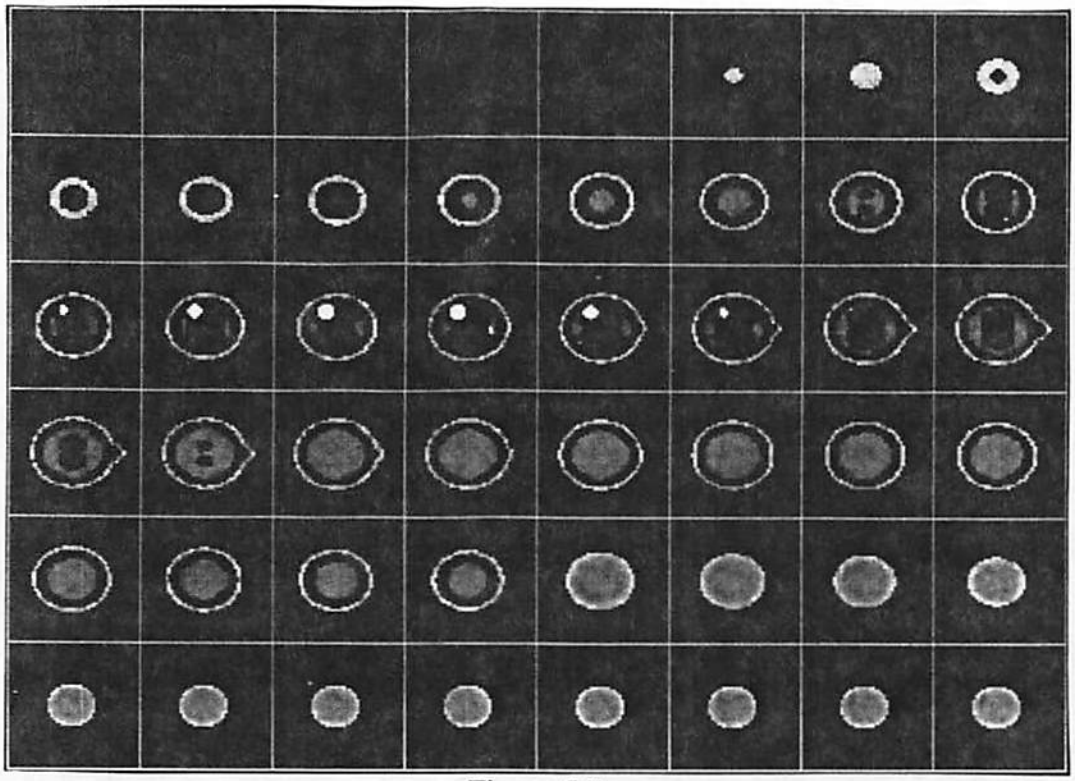

Figure 5.7
MAP reconstruction of the 3-D head phantom in figure 5.2 using Gibbs prior V_2 in equation (5.4).

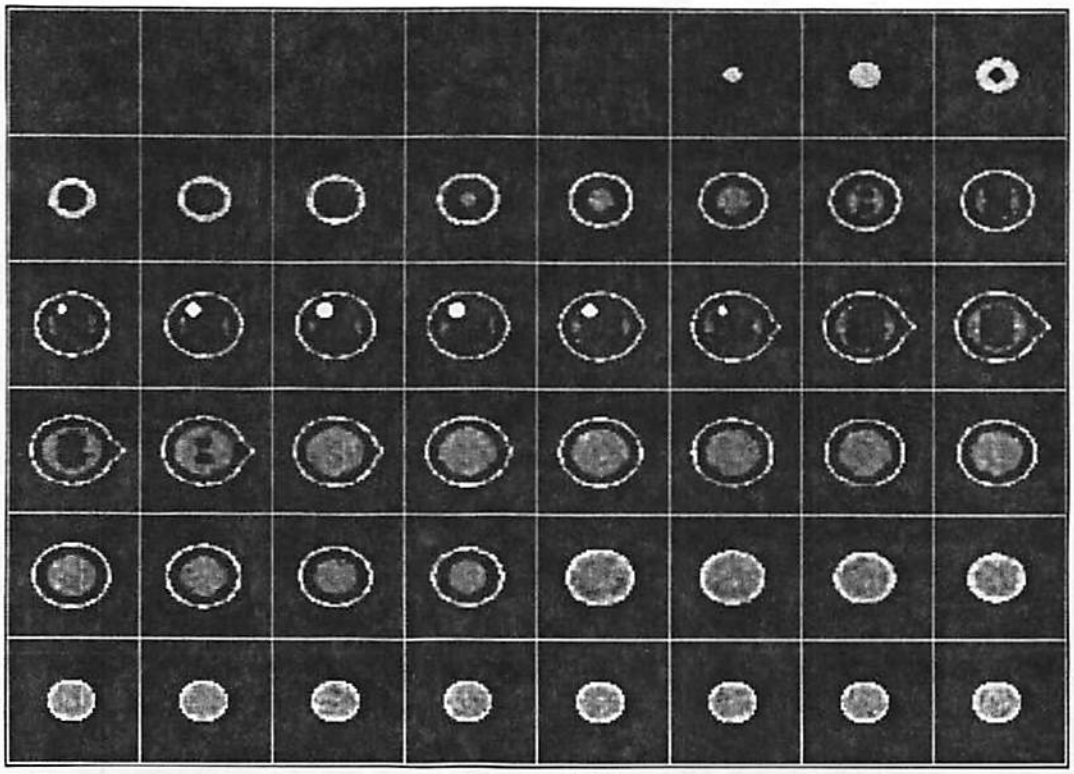

Figure 5.8
MAP reconstruction of the 3-D head phantom in figure 5.2 using Gibbs prior V_3 in equation (5.5).

medical applications where a single pixel incorrectly showing as a bright point within a region of moderate intensity can lead to mis-diagnosis.

The following graphs provide a comparison of the reconstruction approaches based upon the reconstruction error. Herein lies the value of simulation in that the true image is known and can be used in an assessment of performance. Figure 5.9 shows the L_2 error between the true source image and the reconstruction for 100 iterations of the EM ML algorithm and for 100 iterations of the the GEM MAP algorithm using the three different Gibbs priors.

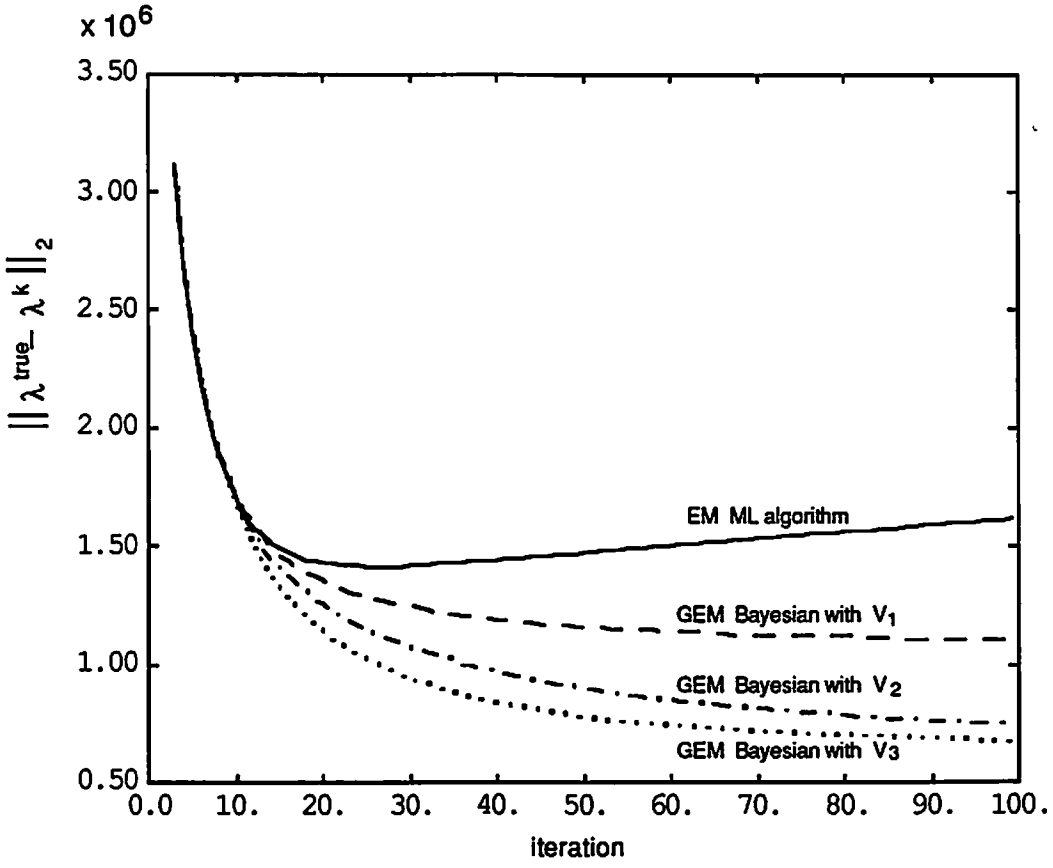

Figure 5.9
Plot of L_2 error between the MAP and ML reconstructions versus iterations of the EM ML and GEM MAP algorithms.

The GEM MAP algorithm initially reduced the L_2 error as quickly as the EM ML algorithm. The EM ML algorithm characteristically iterated away from the true source image after a number of iterations while the GEM MAP algorithm continued to reduce the L_2 error until convergence. The maximum likelihood reconstruction in figure 5.5 which was used for a visual comparison of the reconstruction approaches was produced by halting the EM ML after only 50 iterations. An optimal number of iteration for a ML algorithm can be estimated using the approach in chapter 4. According to the L_2 error criterion, we found an improvement in the reconstruction over the maximum likelihood reconstruction for all three normalized potential functions and all values of $\beta \geq 0.9$. As β was set at values smaller than this, the overinfluence of the prior and underinfluence of the data resulted in reconstructions which were overly smooth. This is shown further in chapter 6.

Figure 5.10 shows the Bayes (likelihood) value at each iteration. This graph shows the monotonic increase in these functions for both the EM ML algorithm and the GEM MAP algorithm and demonstrates simply that these algorithms are performing their function.

These simulations were run in Fortran code on a Sun 3/110 workstation with a floating point accelerator. For the GEM MAP algorithms, a significant improvement in the iteration speed is achieved by setting the allowable parameter space to $\lambda_j \geq \phi > 0$. where ϕ was chosen as some small value such as $\phi = .0001$. If λ_j equaled ϕ and the directional derivative was negative along that coordinate, λ_j was left equal to ϕ and the next pixel was visited. This reduces some of the time spent computing small steps for small pixel values which are converging towards zero, a task particularily time consuming as the algorithm converges. For the 3-D reconstructions, the EM ML algorithm required 48 sec per 3-D iteration while the GEM

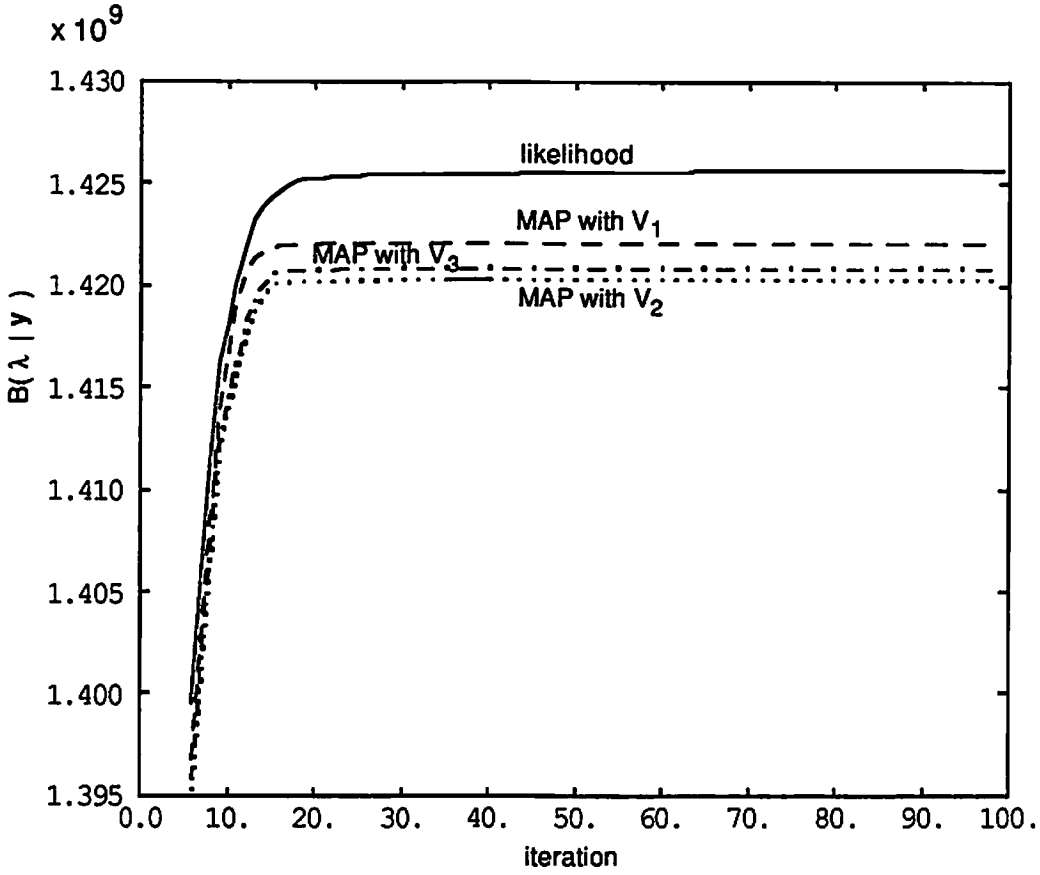

Figure 5.10
The Bayes function value and likelihood versus iterations of the EM ML and GEM MAP algorithms.

MAP algorithm averaged 63 sec per 3-D iteration. The 3-D forward/back projections required the majority of the CPU time (47 sec per iteration) while pixel updating filled the remaining seconds. This computation time could be reduced significantly through the use of an array processor. More importantly, a majority of the computation for these algorithms can be performed in parallel so that reduction of the computation time by a factor roughly equal to the number of viewing angles is entirely possible at this time.

5.6 Conclusion

The GEM MAP algorithm we have presented can be used with any locally correlated priors in the form of Gibbs functions. This algorithm combined the complete/incomplete data formulation of the EM approach, section 3.4.1.2, coordinate ascent, the optimal step-size of a maximum likelihood algorithm derived along parallel lines, and a method of decreasing the step size if necessary. By this approach we were able formulate an algorithm which updates a single image pixel at a time. This approach decoupled some of the interation between the data and image pixels which are not neighbors of the image pixel being updated. In addition, the computational burden of performing a line search was surmounted. This algorithm has the desireable theoretical convergence of its generalized EM formulation. For 3-D images, such as those encountered in emission tomography, this algorithm can provide an improvement over maximum likelihood reconstruction at a nominal increase in computational cost. This work shows that some improvement can be achieved for a wide range of β values, but statistical methods for optimizing the choice of this parameter are still needed. Such a statistical approach is presented in the following chapter.

Chapter 6

Statistical Selection of the Gibbs Prior Parameter

6.1 Introduction

An image model with local spatial dependence which is not restricted or biased in any direction has intrinsic appeal in image processing applications. As discussed in section 3.3, this property is satisfied by non-causal Markov random field models. The Hammersly-Clifford theorem [6] established that the probability distribution characterizing any Markov random field can be expressed as a Gibbs functions by an appropriate choice of potential function. A Gibbs prior can be defined outside of the normalizing constant by specifying a suitable pixel neighborhood, specifying the potential functions on the cliques associated with that neighborhood, and selecting the Gibbs prior parameter. The Gibbs prior parameter specifies the degree to which the maxima of the function stand out. As this parameter varies from 0 to $+\infty$ a MAP restoration or reconstruction varies from totally discounting all data to equivalence with the maximum likelihood solution which depends solely upon the data.

The ease with which non-causal Markov random fields can be characterized through specification of a Gibbs function and the ease with which potential functions may be defined to incorporate image qualities of interest led to the investigation of three Gibbs priors in the previous chapter. In this chapter, we present a statistical approach to selecting the Gibbs prior parameter β. This approach accepts the minimum influence of the prior on the restoration or reconstruction such that a statistic, which is a function of the image estimate and the data, falls within an appropriate confidence interval. This procedure calls for updating of the Gibbs prior

parameter as the iterative MAP reconstruction proceeds. We derive an algorithm to solve the problem as posed and we present a Monte Carlo study of this approach using blurred photon-limited images.

6.2 The Need for a Method of Parameter Selection

As discussed in section 3.2, a Gibbs distribution is a probability measure on the set of configurations $\{\lambda\}$ which has the form

$$f(\lambda) = \frac{1}{K_\beta} e^{\frac{-U(\lambda)}{\beta}} \tag{6.1}$$

where β is the Gibbs prior parameter, K_β is the normalizing constant (partition function), and $U(\lambda)$ is termed the energy function. The normalizing constant K_β is a function of the prior paramenter β. For the discrete case,

$$K_\beta = \sum_{\lambda} e^{\frac{-U(\lambda)}{\beta}} \tag{6.2}$$

and for the continuous case

$$K_\beta = \int e^{\frac{-U(\lambda)}{\beta}} \, d\lambda \tag{6.3}$$

The energy function has the form

$$U(\lambda) = \sum_{c \in C} \sum_{I_c} V_c(\lambda_i .. \lambda_m : \{i..m\} \varepsilon I_c) \tag{6.4}$$

where C denotes the set of all cliques types, I_c is a set of lattice sites in a clique of type c, and $V_c(\lambda)$, termed a potential function, is a function of the values at the sites in I_c.

From the viewpoint of the Gibbs prior, the parameter β simply specifies the degree to which the modes of the prior stand out. As $\beta \rightarrow 0$ the prior becomes increasingly more pronounced about its maxima. As $\beta \rightarrow +\infty$ the Gibbs function tends to a uniform distribution.

MAP estimation of λ using a Gibbs prior presents the following optimization problem

$$\max_{\lambda} \; B(\lambda \mid y;\beta) = \log f(y \mid \lambda) - \frac{1}{\beta} U(\lambda) \tag{6.5}$$

where the normalizing constant K_β has been omitted since it does not affect the location of the maxima. As we select smaller β's, $\beta \rightarrow 0$, the location of the maximum of $B(\lambda \mid y;\beta)$ becomes less dependent upon the data in that the effect of $\log f(y \mid \lambda)$ decreases till the MAP estimate eventually discounts the data. As we select larger β's, $\beta \rightarrow +\infty$, Bayesian estimation becomes unaffected by the prior and MAP estimation reduces to maximum likelihood.

We can examine the effect of different choices for β on an image restoration or reconstruction by examining, for instance, the L_1 error or the L_2 error between the reconstructed image and the true image. In the previous chapter, we presented some simulation results to exhibit the performance of the GEM MAP algorithms presented in that chapter. Using the same computer generated 3-D phantom, the same emission tomography model, and the first GEM MAP algorithm presented in that chapter, we can examine the effect on the MAP reconstruction process using different β values. Figure 6.1 shows the L_2 error between the MAP reconstruction and the true 3-D computer generated image, shown in figure 5.2, for 100 iterations of the GEM MAP algorithm with V_1 of equation (5.3) and four different prior parameter (β) values. The values range from 0.1×180^2 to $+\infty$. Values of β greater

than 0.9×180^2 produced a MAP reconstruction with a lower L_2 error than that in the maximum likelihood recontruction. Values of β less than 0.9×180^2 increased the L_2 due to an overinfluence by the prior.

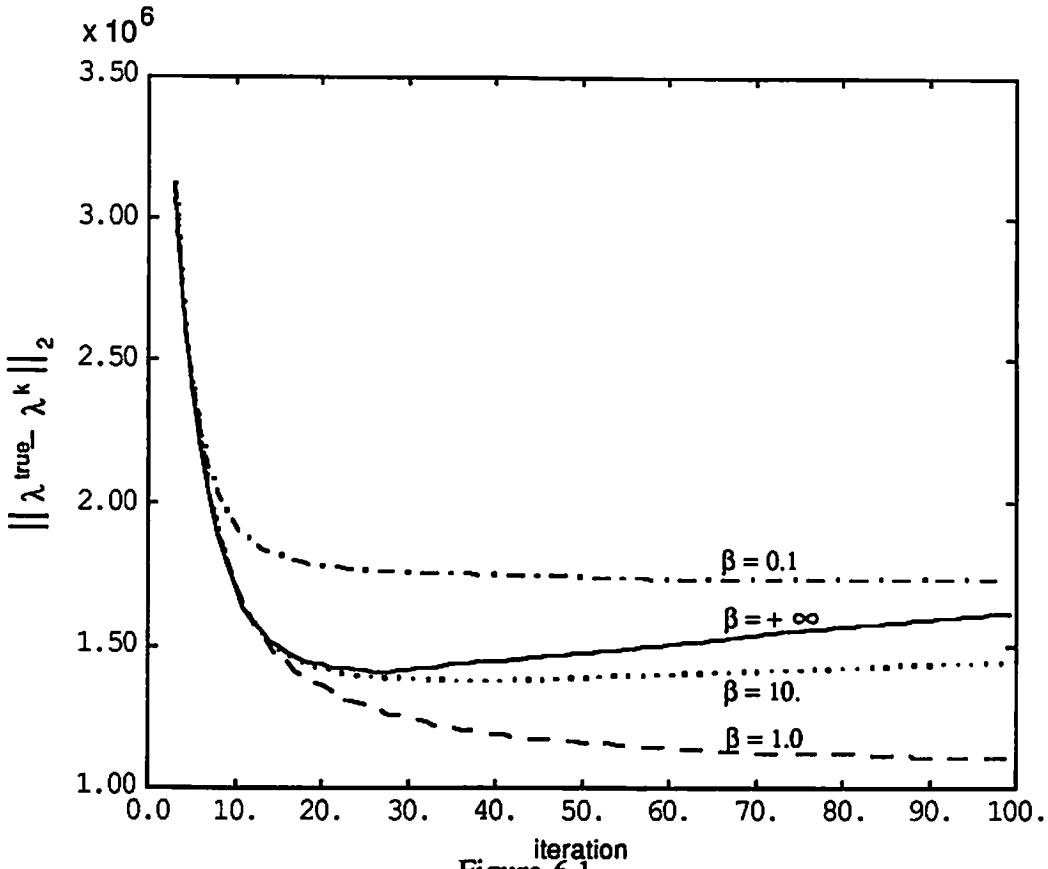

Figure 6.1
Plot of the L_2 error between the MAP reconstruction and the true 3-D image in figure 5.2 for 100 iterations of the GEM MAP algorithm with V_1 and different prior parameter (β) values.

In the preceding chapter, we used trial and error to suitably normalize each of the three Gibbs priors examined in that chapter (figure 5.1) so that the β values we examined would be equally appropriate for all three priors. Using this approach, we compared the performance of the three priors. However, the need for a method of selecting an appropriate β value for MAP restoration or reconstruction of a

degraded image is made apparent by the first three of the following four results:

[1] The same β value is not generally appropriate when different Gibbs priors are used in MAP restorations. While $\beta = 1.0$ may produce good results with one Gibbs prior, $\beta = 32{,}000.0$ may produce comparable results with another.

[2] If we constrain our MAP restorations to use the same Gibbs prior, then the same β value may not be appropriate for different images, even those with the same blurring and noise. When the image histograms are dramatically different or when the images contain different structure, the effect of the prior with the same β value may differ drastically. Better results are obtained using different β values for different images.

[3] If we examine MAP restorations of the same image under different blurring functions and different SNRs, we find that best results are obtained using slightly different β values for each application.

[4] If we examine MAP restorations of the same image with the same blurring and SNR and we use the same Gibbs prior but different realizations of the noise we find that the best results are obtained in each case by essentially the same value of the Gibbs prior parameter.

Although in some applications of MAP estimation with a Gibbs prior an investigator may be able to choose an appropriate β value through trial and error, this is not in general true. With a trial and error method the results reflect the predispositions of the investigator. We present a method of selecting the value for β which imposes a statistic based constraint on the optimization problem of equation (6.5).

Before presenting this method we note that in the context of Bayesian estimation, the regularization parameter γ is equivalent to $1/\beta$, the inverse of the Gibbs prior parameter. Therefore it is informative to examine approaches to selection of

the regularizing parameter which have been proposed in the literature.

6.3 Parameter Selection Methods in Regularization

Image restoration problems are often posed in terms of optimization of a single function $h(\mathbf{y},\boldsymbol{\lambda})$ whose arguments are the data $\mathbf{y}$ and the image estimate $\boldsymbol{\lambda}$. Regularization of the problem can be performed by appending an additional measure $g(\hat{\boldsymbol{\lambda}})$ which is solely a function of the reconstructed image $\hat{\boldsymbol{\lambda}}$. The regularized problem takes the form

$$\max_{\boldsymbol{\lambda}} \; h(\mathbf{y},\boldsymbol{\lambda}) + \gamma g(\boldsymbol{\lambda}) \tag{6.6}$$

where γ is chosen as some value in the range $0<\gamma<+\infty$. As $\gamma \rightarrow 0$ the solution becomes entirely dependent on the first measure $h(\mathbf{y},\boldsymbol{\lambda})$ and as $\gamma \rightarrow +\infty$ the solution becomes independent of the data $\mathbf{y}$. Values of γ between zero and infinity represent trade-offs between the solutions corresponding to those values.

Many papers have been written on the subject of selecting a good regularization parameter. These generally fall into one of four approaches: defining a hyperprior density for the regularizing parameter γ, the method of cross-validation, using a maximum likelihood estimate of γ, and choice of γ based on a fit of the estimate to the data. The first method presents a furthur complication in that an additional basis for the choice of a prior for the regularizing parameter is required. Lacking such a basis in the problem of selecting a Gibbs prior parameter β, we examine the other approaches, all of which were published as applying to the linear data model with i.i.d. Gaussian noise.

For selection of the regularizing parameter by cross-validation [15],[82] based on the linear model $\mathbf{y} = \mathbf{P}\boldsymbol{\lambda} + \mathbf{n}$ where $\mathbf{n}$ is zero-mean i.i.d. Gaussian noise, it is

assumed, as would usually be the case, that the variance of the noise is not known. This approach was posed in the terminology of ridge estimators (section 3.2.2) which are biased estimators of $\hat{\boldsymbol{\lambda}}$ with the form

$$\hat{\boldsymbol{\lambda}} = (\mathbf{P}^t\mathbf{P} + \gamma\mathbf{I})^{-1}\mathbf{P}^t\mathbf{y}. \tag{6.7}$$

Such estimators can be shown to have lower mean squared error over a range of γ values [57],[74]. However, this range is dependent upon the noise variance which is unknown. Cross-validation does not require estimation of the noise variance and is based upon the following notion. Let $\mathbf{P}_{-k}$ be the matrix $\mathbf{P}$ with the k^{th} row omitted. Set aside the k^{th} data value y_k and form the regularized solution based upon the remaining data

$$\hat{\boldsymbol{\lambda}}_{-k} = (\mathbf{P}_{-k}^t\mathbf{P}_{-k} + \gamma\mathbf{I})^{-1}\mathbf{P}_{-k}^t\mathbf{y} \tag{6.8}$$

Then, the cross-validation idea is that the estimate of the omitted data value given the regularized solution should be close to the data value, i.e. $| y_k - \sum_j \mathbf{P}(k,j)\hat{\boldsymbol{\lambda}}_{-k}(j) |^2$ should be small. Ordinary cross-validation selects γ as the minimizer of

$$\frac{1}{n}\sum_{k=1}^{n}[y_k - \sum_j \mathbf{P}(k,j)\hat{\boldsymbol{\lambda}}_{-k}(j)]^2 \tag{6.9}$$

The generalized cross-validation method selects γ to minimize a particular weighted version of equation (6.5). It can be shown [26] that a γ chosen by this method is a biased estimate of the γ minimizing $\frac{1}{n}E\{\|\mathbf{P}\boldsymbol{\lambda} - \mathbf{P}\hat{\boldsymbol{\lambda}}\|\}$ where $\hat{\boldsymbol{\lambda}}$ is given in equation (6.7). Unlike ordinary cross-validation, the generalized cross-validation function always has a unique minimum [26]. In Monte Carlo studies [26] generalized cross-validation has been shown to perform well in comparison to other parameter selection methods. However, its application is limited to the linear

model. What may be worse is the difficulty in minimizing equation (6.9) due to the large dimension of $\boldsymbol{\lambda}$ in image processing applications.

Selection of γ by the method of maximum likelihood [17],[26] is also based on the linear model $\mathbf{y} = \mathbf{P}\boldsymbol{\lambda} + \mathbf{n}$ where $\mathbf{y}$ is the vector of data, $\mathbf{P}$ is a known linear transformation, and $\mathbf{n}$ is $N(0,\sigma^2\mathbf{I})$. If we assume that the elements of unknown vector $\boldsymbol{\lambda}$ are also normal $N(0,a^2\mathbf{I})$, then the data $\mathbf{y}$ is $N(0,a^2\mathbf{P}\mathbf{P}^t + \sigma^2\mathbf{I})$ and the posterior mean and MAP estimate have the form of a regularized solution, or equivalently a ridge regression.

$$\hat{\boldsymbol{\lambda}} = (\mathbf{P}^t\mathbf{P} + \gamma\mathbf{I})^{-1}\mathbf{P}^t\mathbf{y} \tag{6.10}$$

where $\gamma = \dfrac{\sigma^2}{a^2}$. In addition, the likelihood function for the regularization or ridge parameter γ, can be formed [26] so that a maximum likelihood estimate of γ can be used in the standard ridge regression form, equation (6.7). Computation of the ML selection of γ requires minimization of $M(\gamma)$ where

$$M(\gamma) = \frac{1}{N}\,\frac{\mathbf{y}^t(I - A(\gamma))\mathbf{y}}{|I - A(\gamma)|^{1/N}} \tag{6.11}$$

and $A(\gamma) = \mathbf{P}(\mathbf{P}^t\mathbf{P} + \gamma\mathbf{I})^{-1}\mathbf{P}^t$ [26]. This can be quite difficult for large N, the dimension of the data vector, due in part to the determinant term in the denominator. An implementation of this approach in the Fourier domain [17] seems to be computationally easier. But, this approach has limited application due to the prior assumption that the elements of the vector $\boldsymbol{\lambda}$ are i.i.d. normal, which is generally not reasonable. However, by subtracting a simple estimate of the local mean of the image, this Guassian assumption has been successfully applied [45].

Another approach which has been applied is choice of the regularizing parameter based on a fit of the resulting estimate to the data [75],[76],[77]. It has long

been recognized that for the linear Gaussian model the errors $y_i - (\mathbf{P}\lambda)_i$ are i.i.d. normal so that the squared errors $\| \mathbf{y} - \mathbf{P}\hat{\lambda} \|_2^2$ have a chi-squared distribution with N degrees of freedom. Therefore, one approach to the selection of the regularizing parameter in equation (6.8) has been to choose γ such that the regularized solution solution $\hat{\lambda}$ satisfies $\| \mathbf{y} - \mathbf{P}\hat{\lambda} \|_2^2 = N$ or $N+2\sqrt{2N}$ the mean or 95th percentile of a chi-squared distribution of N degrees of freedom. Additional statistics such as the Kolmogorov-Smirnov statistic [52] have also been suggested to test the normality of the errors. The method for selection of the Gibbs prior parameter presented in this chapter may be considered as being in the same spirit in that a statistic which is a function of the restored image is used to determine a suitable parameter value.

6.4 Parameter Selection Methods in MAP

Two methods of solving the problem addressed in this section have been proposed by other authors. The first was proposed by J. Meritus "Self-Calibrating Bayeian Methods for Image Reconstruction in Emission Tomography," Ph.D. thesis, Brown University (1987). Although I have not been able to obtain a copy of this thesis at the time of completion of this manuscript, the approach was described in minor detail by Geman and McClure in [23]. It uses the complete/incomplete data formulation of Dempster, Laird, and Rubin to form an EM algorithm. The complete data is taken to be the unknown image λ while the incomplete data is the observed data vector $\mathbf{y}$. In [21] the computational demands of this approach are deemed enormous but not prohibitive.

In [23], Geman and McClure devise a method for estimating the Gibbs prior parameter β directly from the data without having to perform an intermediate reconstruction of the unknown image in emission tomography. In order to clearly

explain their approach, it is necessary to cast the problem in the emission tomography framework. Let the data consist of numbers of emissions of gamma rays detected by a gamma camera. The camera collects a set of data at each of N_a viewing angles. The data collected by the camera at any angle can be represented by a 2-D image, each pixel corresponds to a pixel region on the face of the camera. Let us store each such data image in a matrix, so that matrix $\mathbf{Y}^k$ contains the 2-D data image collected at viewing angle k of N_a different viewing angles. Then, proceeding as in section 2.3.2.3, the mean of the 2-D data image $\mathbf{Y}^k$ is equal to a projection $\mathbf{P}^k$ of the mean of the 3-D isotope source distribution stored as a matrix Λ, i.e. $E\{\mathbf{Y}^k\} = \mathbf{P}^k\Lambda$. Then the statistic $M(\mathbf{Y}^1 \cdots \mathbf{Y}^{N_a})$ which Geman and McClure examine is a measure of roughness of the data since its main components measures a squared difference between neighboring data values. Let matrix $\mathbf{Y}^k$ be indexed in matrix fashion so that $\mathbf{Y}^k_{ij}$ denotes the i^{th} row and j^{th} column. Then

$$M(\mathbf{Y}^1 \cdots \mathbf{Y}^{N_a}) = \sum_k \sum_i \sum_j \left[\left(\frac{\mathbf{Y}^k_{i,j}}{\sum_q \mathbf{P}^k_{j,q}} - \frac{\mathbf{Y}^k_{i,j+1}}{\sum_q \mathbf{P}^k_{j+1,q}} \right)^2 - \frac{\mathbf{Y}^k_{i,j}}{\left(\sum_q \mathbf{P}^k_{j,q}\right)^2} - \frac{\mathbf{Y}^k_{i,j+1}}{\left(\sum_q \mathbf{P}^k_{j+1,q}\right)^2} \right] \quad (6.12)$$

For small values of this measure, the isotope distribution should be better represented as having been generated from a Gibbs distribution with a small β value, i.e. from a distribution which places dramatically more probability upon smooth images. For large values of this measure, the isotope distribution should be better represented as having been generated from a Gibbs distribution with a large β value, i.e. from a distribution does not place such large probability upon smooth images. The expected value of $M(\mathbf{Y}^1 \cdots \mathbf{Y}^{N_a})$ is

$$E\{M(\mathbf{Y}^1 \cdots \mathbf{Y}^{N_a})\} = \sum_k \sum_i \sum_j \left[\frac{(\mathbf{P}^k\Lambda)_{i,j}}{\sum_q \mathbf{P}^k_{j,q}} - \frac{(\mathbf{P}^k\Lambda)_{i,j+1}}{\sum_q \mathbf{P}^k_{j+1,q}} \right]^2 \quad (6.13)$$

In order to use this statistic to estimate the parameter β, they generated an estimate of its mean value for a range of β values. To do this, they generated 230 configurations Λ from the Gibbs prior using 46 different β values, 5 configurations per β value. Given each Λ, they computed the expected value in equation (6.13). Then they formed an estimate of $E_\beta\{M(\mathbf{Y}^1 \cdots \mathbf{Y}^{N_a})\}$ by fitting a cubic spline regression to these computed values. Now, given a set of data $\{\mathbf{Y}^1 \cdots \mathbf{Y}^{N_a})\}$ from the emission tomography system modeled by the system matrices $\{\mathbf{P}^1..\mathbf{P}^{N_a}\}$ used in the computation of the estimate $\hat{E}_\beta\{M(\mathbf{Y}^1 \cdots \mathbf{Y}^{N_a})\}$, the statistic $M(\mathbf{Y}^1 \cdots \mathbf{Y}^{N_a})$, is computed. The parameter value β is chosen as that for which

$$M(\mathbf{Y}^1 \cdots \mathbf{Y}^{N_a}) = \hat{E}_\beta\{M(\mathbf{Y}^1 \cdots \mathbf{Y}^{N_a})\} \tag{6.14}$$

It is very desireable to use a statistic which is solely a function of the data. Geman and McClures approach was documented in [23] and several reconstructions were presented. These reconstructions should be compared to similar reconstructions using different β values so that the results of this approach can be fully evaluated.

Given a set of data, computation of $M(\mathbf{Y}^1..\mathbf{Y}^{N_a})$ requires only a modest amount of computation. Computation of $\hat{E}_\beta\{M(\mathbf{Y}^1 \cdots \mathbf{Y}^{N_a})\}$ by Geman and McClure required 41 CPU hours on a 100 Megaflop array processor. If $\hat{E}_\beta\{M(\mathbf{Y}^1 \cdots \mathbf{Y}^{N_a})\}$ has been computed and stored, selection of the parameter β only requires equating $M(\mathbf{Y}^1 \cdots \mathbf{Y}^{N_a})$ to a point on the curve. However, this curve $\hat{E}_\beta\{M(\mathbf{Y}^1 \cdots \mathbf{Y}^{N_a})\}$ only pertains to a single Gibbs prior. The prior used by Geman and McClure, equation (5.4), has an additional parameter δ which is fixed based on the range of values in the unknown image. For longer data collection times or larger isotope doses the total counts will be higher so that the range of values in the unknown image will change. The parameter δ would need to be increased accordingly. Similarly, the statistic $M(\mathbf{Y}^1 \cdots \mathbf{Y}^{N_a})$ will have a higher

mean value. The curve $\hat{E}_\beta\{M(\mathbf{Y}^1 \cdots \mathbf{Y}^{N_a})\}$ would need to reflect this. The result may be that the curve $\hat{E}_\beta\{M(\mathbf{Y}^1 \cdots \mathbf{Y}^{N_a})\}$ needs to be computed for different mean values of the isotope distribution and for different δ values.

Additionally, $\hat{E}_\beta\{M(\mathbf{Y}^1 \cdots \mathbf{Y}^{N_a})\}$ pertains only to the particular system model represented by the matrices $\{\mathbf{P}^1..\mathbf{P}^{N_a}\}$. In describing a real emission imaging system, these matrices are dependent on the day-to-day record of the calibration of the machine. This calibration is typically recorded as a 2-D flood field image. In addition, different energy isotopes result in different point spread functions for a given collimator. Again, the result may be that the curve $\hat{E}_\beta\{M(\mathbf{Y}^1 \cdots \mathbf{Y}^{N_a})\}$ needs to be computed for different flood field calibration images, different emmision imaging systems, and for isotopes of different energies.

The computational demands of estimating $\hat{E}_\beta\{M(\mathbf{Y}^1 \cdots \mathbf{Y}^{N_a})\}$ are sufficiently formidable, 41 CPU hours on a 100 Megaflop array processor, that the necessity of computing a family of these curves due to the reasons presented above may make this approach less practical.

Finally, although the performance of this approach needs to be evaluated more fully, it offers one statistical approach to selecting a Gibbs parameter value β. In the next section we offer an alternative statistical approach. and we evaluate the performance of that approach against the complete range of potential β values.

6.5 Statistic Based Selection of the Gibbs Prior Parameter

We formulate the selection of β by requiring the MAP estimate to be minimally influenced by the prior while being consistent with the data as evaluated by a statistic. Let λ_β^k be an estimate of the true image after the k^{th} iteration of an iterative algorithm for finding the maximum of $B(\lambda | y;\beta)$ equation (6.5). Choose an

appropriate statistic $s[\lambda_\beta^k;y]$, which is a function of β through the effects of β on the estimate λ_β^k, and choose an appropriate confidence interval $[\gamma_1, \gamma_2]$ for that statistic. Where $\hat{\lambda}_\beta$ signifies the estimate of λ at convergence of a MAP algorithm using Gibbs prior parameter β, a necessary condition for a statistic to be suitable is that it satisfy

$$s[\hat{\lambda}_0;y] > \gamma_2 \quad \text{and} \quad s[\hat{\lambda}_\infty;y] < \gamma_1. \tag{6.15}$$

For the Gaussian and Poisson cases, we will specify simple statistics satisfying this property later. Choose $\hat{\beta}$ and $\hat{\lambda}_\beta$ as solutions to the problem

$$\hat{\beta} = \text{the max } \beta \text{ such that the } \hat{\lambda}_\beta \text{ maximizing } B(\lambda \mid y;\beta) \tag{6.16}$$

$$\text{satisfies } s[\hat{\lambda}_\beta;y] \in [\gamma_1,\gamma_2] \text{ where } B(\lambda \mid y;\hat{\beta}) \text{ is given in equation (6.5).}$$

Let Ψ signify the set $\beta > 0$ and Ω the set $\lambda \geq 0$. For a given data vector y, $s[\lambda;y]$ is a functional defined over the set Ω. Given a data vector y, an initial estimate λ^0 and any specified deterministic MAP algorithm, the set $\{\hat{\lambda}_\beta\} \subset \Omega$ are the MAP estimates generated by the algorithm using Gibbs parameter values $\beta \in \Psi$. This set is then well-defined with a one-to-one mapping from $\{\beta\}$ to $\{\hat{\lambda}_\beta\}$. The mapping from $\{\beta\}$ to $s[\hat{\lambda}_\beta;y]$ can be many-to-one , but in general, we would expect it to be in large part one-to-one. Given y and λ^0, the inequality specifying the confidence interval for the statistic $\gamma_1 < s[\hat{\lambda}_\beta;y] < \gamma_2$ defines the subset of acceptable β values in Ψ. We seek a β such that λ which solves $\max_\lambda B(\lambda \mid y;\beta)$ lies in this subset. Since we want the reconstruction to be minimally influenced by the prior we select the largest β for which this holds. From equation (6.15) with the assumption of continuity of $s[\hat{\lambda};y]$ in the region of the lower bound, we formulate the solution as

$$\hat{\beta} = \text{the max } \beta \text{ such that the } \hat{\lambda} \text{ maximizing } B(\lambda \mid y,\beta) \text{ satisfies } s[\hat{\lambda};y] = \gamma_l. \quad (6.17)$$

where γ_l is the lower limit of the confidence interval of statistic $s[\hat{\lambda};y]$.

6.6 Appropriate Statistics for the Gaussian and Poisson Reconstruction Problems

The Poisson data model assumes the data y to be Poisson distributed with mean equal to a blurred version of the uncorrupted image λ. The uncorrupted image λ is allowed to take the complete range of non-negative values. Let $E\{y\} = P\lambda$ equal the blurred version of the uncorrupted image λ. Then the data y are Poisson with mean $P\lambda$, so that

$$f(y \mid \lambda) = \prod_i e^{-(\sum_j P_{ij}\lambda_j)} \frac{(\sum_j P_{ij}\lambda_j)^{y_i}}{(y_i)!} . \quad (6.18)$$

For this data model, a form of Pearson's chi-squared statistic satisfies the necesary criterion. As shown in appendix 4a, estimates of the parameters of a multivariate Poisson distribution and the data are formulated into Pearson's statistic χ_p^2 as

$$\chi_p^2 = \sum_{m=1}^{M} \frac{[y_m - S_d \pi_m]^2}{S_d \pi_m} \quad (6.19)$$

where $\pi_k = \frac{(P\lambda)_k}{\sum_m (P\lambda)_m}$ and $S_d = \sum_m y_m$, the sum of the data. Pearson showed that as $S_d \to +\infty$ the distribution of χ_p^2 converges to a chi-squared distribution with $M-1$ degrees of freedom. If N of the parameters of the statistic χ_p^2 have been estimated by any asymptotically normal and asymptotically efficient estimator, such as maximum likelihood, the distribution of χ_p^2 is asymptotically chi-squared with

$M-1-N$ degrees of freedom [14]. Let us choose a suitable confidence interval $[\gamma_1, \gamma_2]$ for a chi-squared variable with $M-1$ degrees of freedom. If we use maximum likelihood reconstruction to estimate the N λ_i's and N is of moderate size with respect to $M-1$, then the ML solution $\hat{\lambda}_{ML}$ will invariably result in ${\chi_p}^2 \ll \gamma_1$ with probability approximately equal to one.

Recall that as $\beta \rightarrow +\infty$ the Gibbs prior tends to a uniform distribution and that MAP estimation with a uniform prior is equivalent to MLE. So, as $\beta \rightarrow +\infty$, ${\chi_p}^2 \ll \gamma_1$ with probability ≈ 1. The Gibbs priors defined using the potential functions in equations (5.3),(5.4), and (5.5) are maximized by any uniform image, i.e. any image for which all pixels are equal. As $\beta \rightarrow 0$, the MAP estimate discounts $\log f(\mathbf{y}|\boldsymbol{\lambda})$ in equation (6.3), so that the MAP estimate is a uniform image of constant intensity. This constant image with no structure will yield a statistic value which is above the chosen confidence interval since it does not adhere at all to the data, i.e. ${\chi_p}^2$ evaluated at $\boldsymbol{\lambda}=\boldsymbol{\lambda}_c$ will result in ${\chi_p}^2 \gg \gamma_2$. Therefore, this statistic satisfies the requirements of section 6.4 for a suitable statistic in this application.

For the linear Gaussian model $\mathbf{y} = \mathbf{P}\boldsymbol{\lambda} + \mathbf{n}$ where $\mathbf{y}$ is the data, $\boldsymbol{\lambda}$ is the unknown image, and $\mathbf{n}$ is a vector of i.i.d. Gaussian variables, the squared innovations $\sum_i [\mathbf{y}_i - (\mathbf{P}\boldsymbol{\lambda})_i]^2$ is a statistic which is asymptotically chi-squared distributed with M degrees of freedom. The treatment of this case proceeds identically to that of the Poisson data model so that for $\beta = 0$ and $\beta = +\infty$, this statistic satisfies the two requirements for a suitable statistic in this application.

MAP restoration using both a Gibbs prior and the method of selecting the prior parameter we have presented has another nice property. In Bayesian approaches it is desired that as more and more data are acquired the effect of the data on the solution should eventaully swamp that of the prior. Let the number of

data values be M and the number of parameters or image pixels estimated be N. To satisfy this Bayesian property, as $M \rightarrow +\infty$ we want $\beta \rightarrow +\infty$ so that MAP estimation tends to ML estimation. The selection of the Gibbs prior parameter is based on a confidence interval for a chi-squared variable of $M-1$ degrees of freedom. The statistics discussed in the previous section used have an asymptotic chi-squared distribution of $M-1-N$ when ML estimation is used. Therefore,as the number of data values M increases, $M-1-N \rightarrow M-1$ so that ML estimation should result in a statistic value within the prescribed confidence interval.

6.7 Some Restoration Examples

We have formulated the criterion for selection of the Gibbs prior parameter in the preceeding sections and we have discussed suitable statistics to use for the Poisson and Gaussian data models. In the next section we will present an algorithm for achieving an estimate and a parameter value which satisfy this criterion. At this point we can evaluate the performance of this method with a few test cases, albeit through great computational cost. Given a noisy image, we can perform MAP restorations for different values of the Gibbs prior parameter; recording the statistic value, its lower confidence bound, and the L_1 and L_2 restoration error. The correct β value is easily seen as that for which the statistic is equal to the lower confidence bound. For large N, the lower confidence bound of a chi-squared variable of N degrees of freedom can be computed as

$$L_{\chi^2} = \frac{1}{2}(z_p + \sqrt{2N-1})^2 \tag{6.20}$$

where z_p is the corresponding lower percentile for a two-sided confidence interval of a standard normal variable. Were the Gibbs prior an ideal probabilistic description for an image, one would normally chose a 95% or a 99% confidence interval.

For instance, $z_p = -3.0$ gives a two-sided 99.87% confidence interval. However, images are much more complex so that no prior is completely sufficient. We would therefore, like to decrease the influence of the prior somewhat. This is done by widening the confidence interval. We chose the standard normal percentile $z_p = -12.0$. for use in equation (6.20). This lowers the confidence bound by a small amount so that the influence of the prior is decreased somewhat over that which we would accept from an ideal prior.

For these test cases, we examine restorations of two images. The first is a 256x256 image of the moon and the second is a 256x256 image of a boat. The uncorrupted images were convolved with a 3x3 blurring function and the mean of the first blurred image was set to 35, that of the second was set to 40. These images were used as the mean of a Poisson data image. Figure 6.2 shows the uncorrupted and the noisy images of the moon's surface.

In the previous chapter we examined three Gibbs prior for their use in Bayesian image reconstruction applied to the emission tomography problem. We will examine these same three Gibbs priors for their value as a general purpose image prior in Bayesian restoration of blurred and noisy images. The potential functions of the three priors are defined in equations (5.3),(5.4), and (5.5). We selected a 2^{nd} order neighborhood consisting of the 8 nearest neighbors. The potential function containing pixels diagonally offset from one another was weighted by $\frac{1}{\sqrt{2}}$. A free boundary [21] was used for pixels along the edges of the image. These pixels have fewer neighbors than pixels in the interior of the image. The four corner pixels have three neighbors while pixels along the edges have five.

The ML and MAP restorations based upon this statistical method of selecting the prior parameter values are shown in figures 6.3 and 6.4. The visual quality of

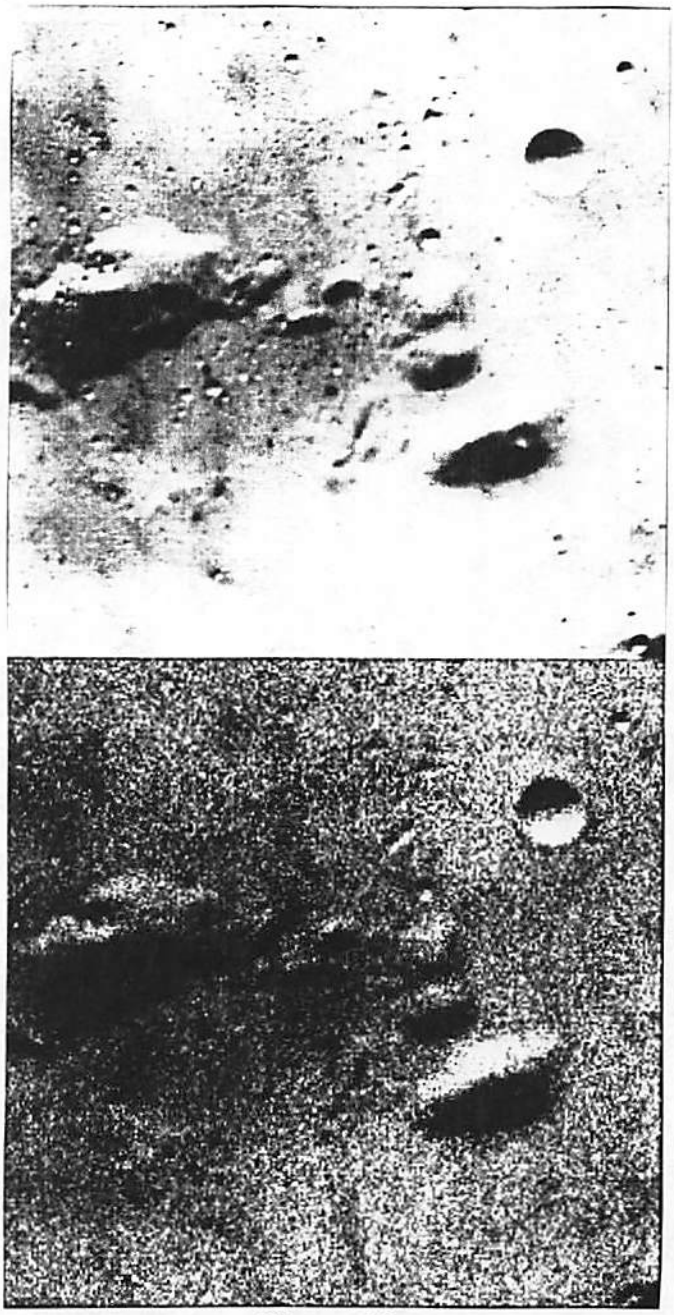

Figure 6.2
The uncorrupted and noisy 256x256 images of the moon's surface. The bottom noisy image was formed by blurring with a 3x3 Gaussian shaped function, normalizing to an average pixel value of 35, and using this as the mean of a 2-D Poisson process.

Figure 6.3
ML and Map restorations of the image in figure 6.2. Top image - ML restoration using the stopping criterion of chapter 4. Bottom image - MAP restoration using potential function V_1 of equation (5.3) and the the statistical method of selecting β.

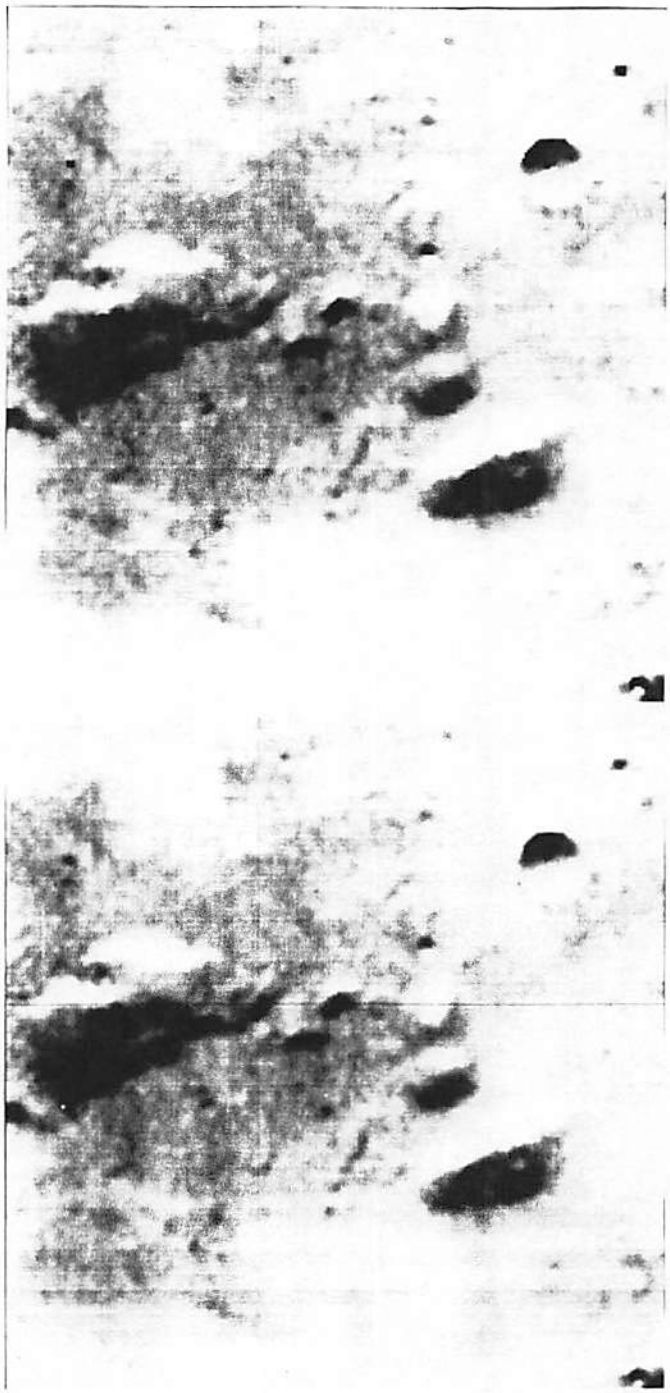

Figure 6.4
Top image - MAP restoration of the image in figure 6.2 using potential function V_2 of equation (5.4) and the statistical method of selecting β. Bottom image - MAP restoration using potential function V_3 of equation (5.5) and the statistical method of selecting β.

the restorations based on V_2 and V_3 are clearly the best two of the four. MAP restoration using V_1 also outperformed ML in this respect. Visual comparison does not tell the whole story so that some measure of restoration error is desired. In addition, this measure of error should be presented for a wide range of possible β values so that some judgement as to how well the parameter selection method is performing can be made. Figures 6.5, 6.6, and 6.7 show the L_1 and L_2 restoration error performance of the MAP restorations of the images of the moon's surface for the three Gibbs potential function, equations (5.3), (5.4), and (5.5) respectively and a wide range of β values. For each MAP image restoration, the β value selected according to the statistical method is easily seen as that for which the statistic is equal to the lower confidence bound in each graph. The accompanying L_1 and L_2 restoration error of the MAP estimate corresponding to that β value can be read from the graph. As these graphs show, the statistical method of selecting an appropriate β value as presented here provides a very promising approach which selects near optimal β values for a variety of Gibbs priors used in MAP restorations.

The undegraded image of the surface of the moon is quite smooth and lacks some of the detail found in other images. As a further comparison, the same restorations are performed using an image of some boats in a harbor; an image possessing dramatically more detail. This image has some additional structure such as the boat masts and clouds which are useful for comparing restorations. Figure 6.8 shows the uncorrupted and the noisy images of the boat. The ML and MAP restorations based upon those parameter values are shown in figure 6.9 and 6.10. Again, some measure of restoration error should be shown for a wide range of possible β values. Figures 6.11, 6.12, and 6.13 show the performance of the MAP

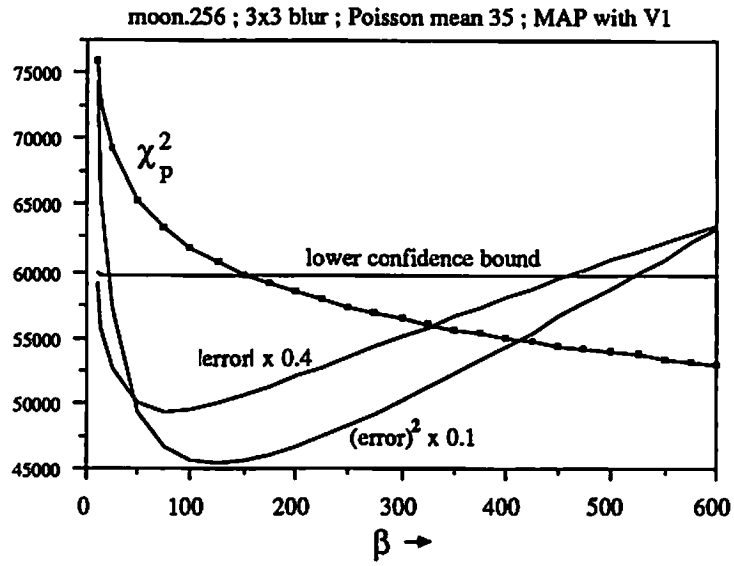

Figure 6.5
Plot of the L_1 and L_2 restoration error versus values of the Gibbs prior parameter in MAP restorations of figure 6.2 using potential function V_1 equation (5.3).

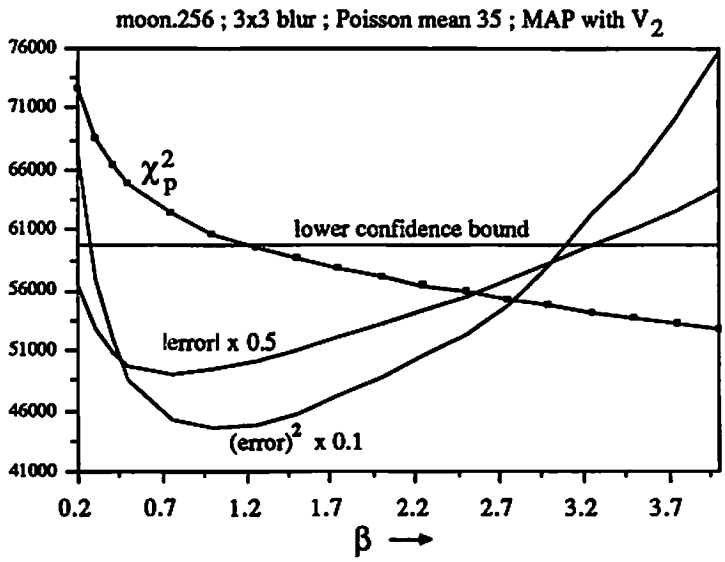

Figure 6.6
Plot of the L_1 and L_2 restoration error versus values of the Gibbs prior parameter in MAP restorations of figure 6.2 using potential function V_2 equation (5.4).

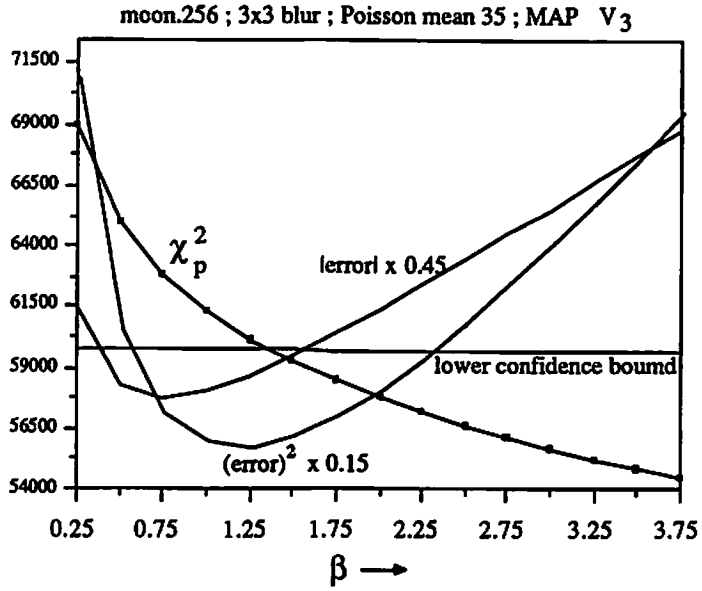

Figure 6.7
Plot of the L_1 and L_2 restoration error versus values of the Gibbs prior parameter in MAP restorations of figure 6.2 using potential function V_3 equation (5.5).

restorations of these images for the same three Gibbs potential function, equations (5.3), (5.4), and (5.5), with respect to L_1 and L_2 restoration error over a wide range of β values. Again, the Gibbs prior parameter value as selected by the statistic based method of the preceeding section is easily determined from the figures. As these examples have shown, the statistical method presented in this chapter has been shown to be capable of selecting β values reasonably near the optimal one for a variety of Gibbs priors used in MAP restorations of widely varying images. The selected β values have varied with both the Gibbs prior used and the particular image restored. If we go back to figures 6.5 and 6.11 and compare appropriate β values for MAP restoration based on the Gibbs prior with potential function V_1 of equation (5.3) this becomes apparent. The image of the moons surface possesses a high degree of smoothness when compared to the image of the boats. Therefore

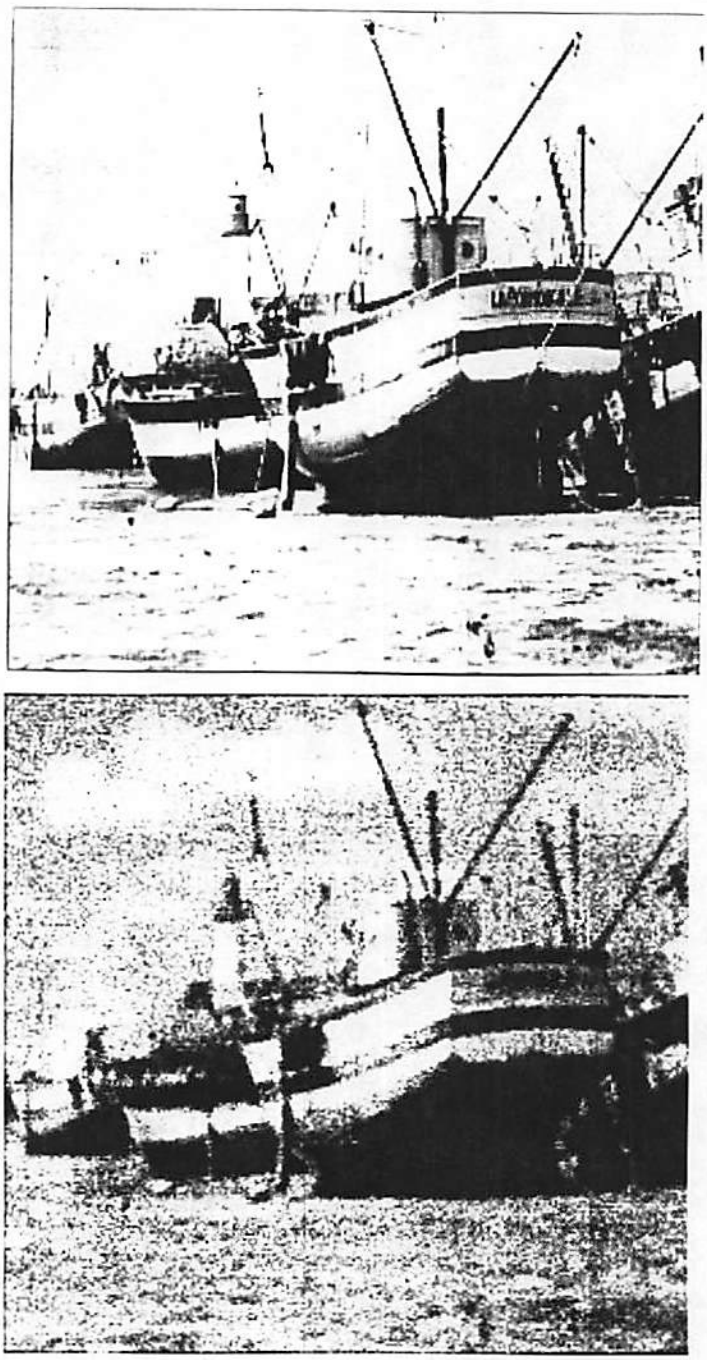

Figure 6.8
The uncorrupted and noisy 256x256 images of the boat. The bottom noisy image was formed by blurring with a 3x3 Gaussian shaped function, normalizing to an average pixel value of 40, and using this as the mean of a pseudo-random 2-D Poisson process.

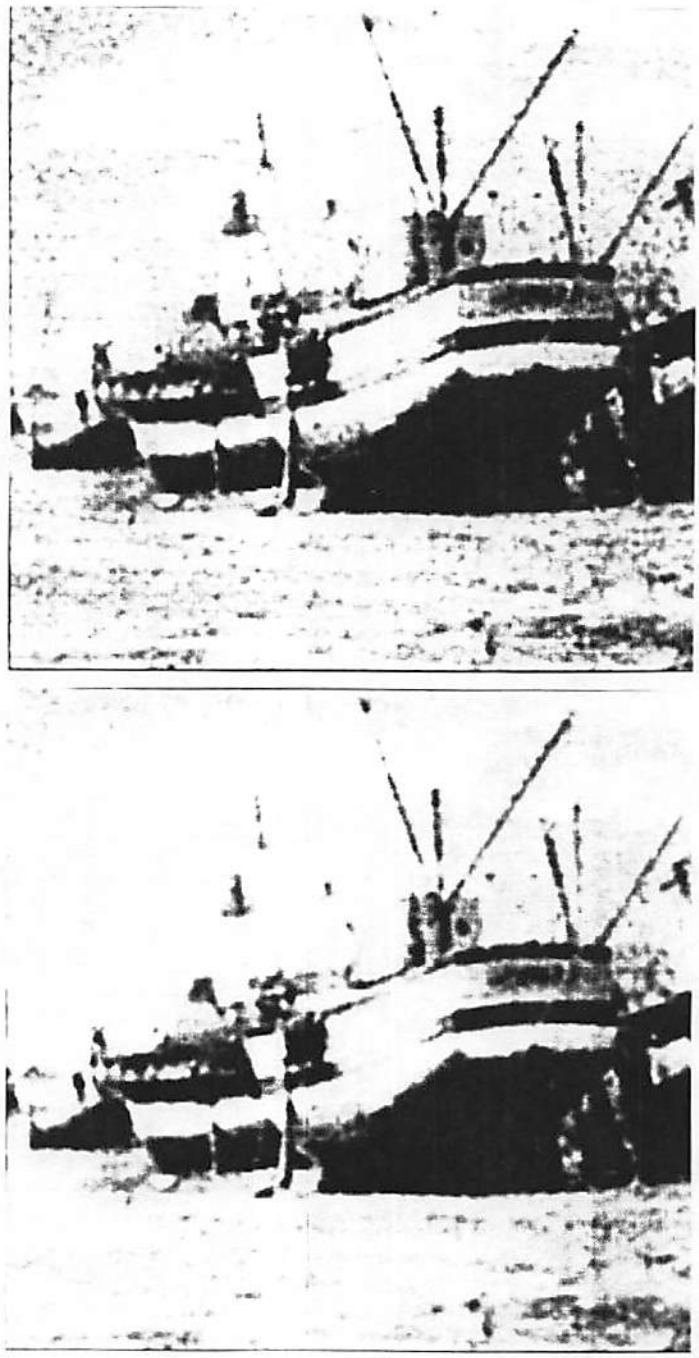

Figure 6.9
ML and Map restorations of the image of a boat in figure 6.6. Top image - ML restoration using the stopping criterion of chapter 4. Bottom image - MAP restoration using potential function V_1 of equation (5.3) and the the statistical method of selecting β.

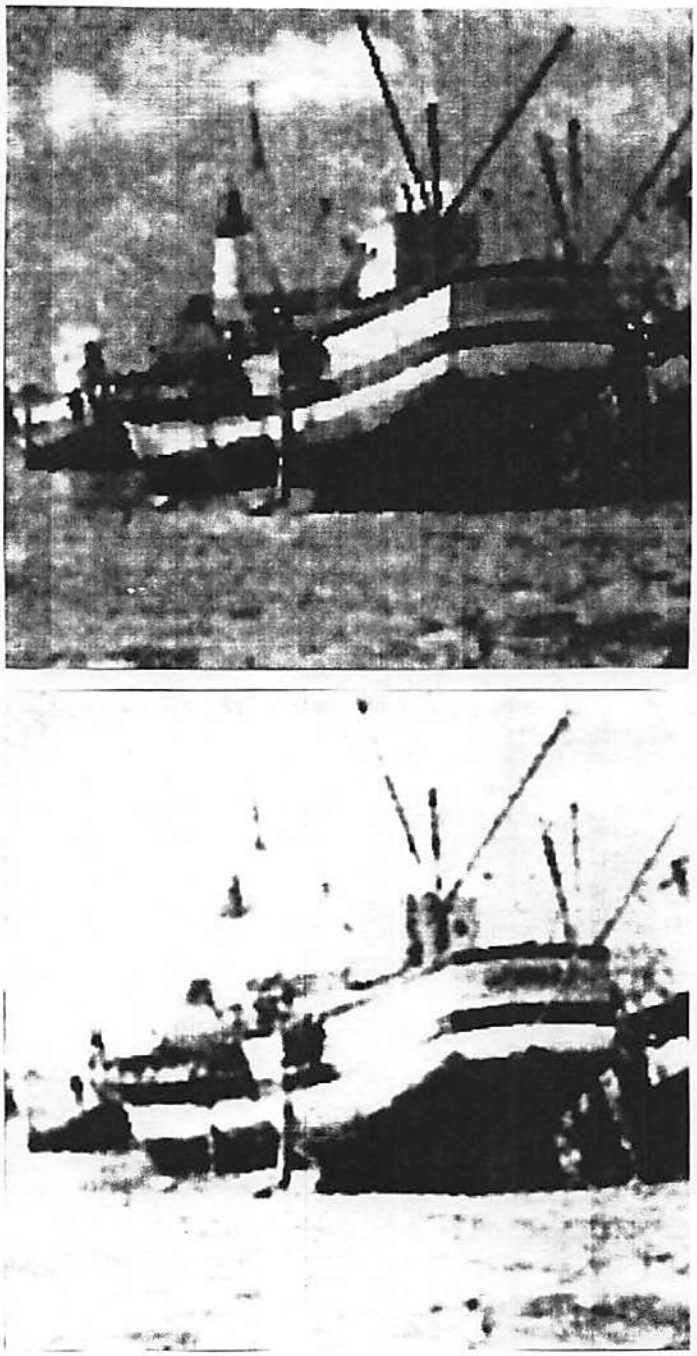

Figure 6.10
Map restorations of the image of a boat in figure 6.6. Top image - using potential function V_2 of equation (5.4) and the statistical method of selecting β. Bottom image - using potential function V_3 of equation (5.5) and the statistical method of selecting β.

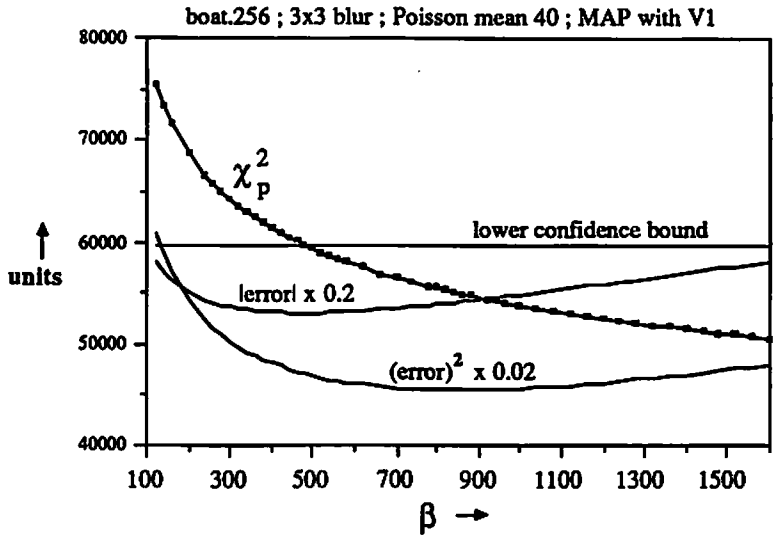

Figure 6.11
Plot of the L_1 and L_2 restoration error versus values of the Gibbs prior parameter in MAP restorations of figure 6.6 using potential function V_1 equation (5.3).

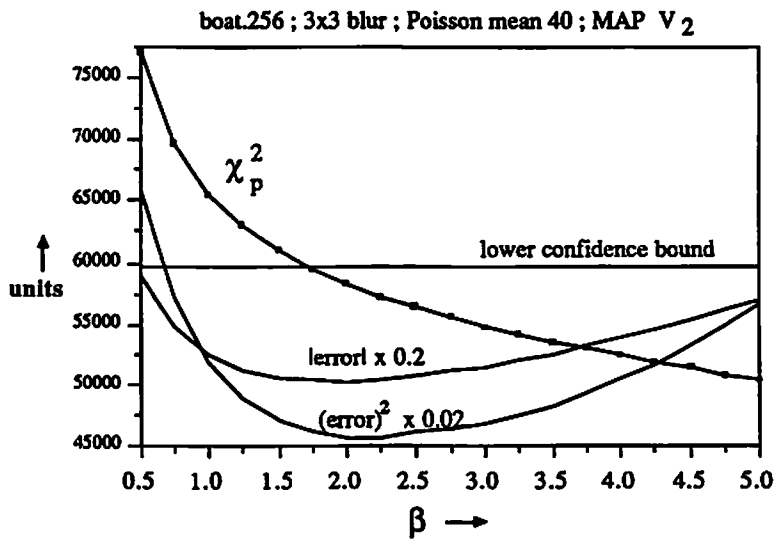

Figure 6.12
Plot of the L_1 and L_2 restoration error versus values of the Gibbs prior parameter in MAP restorations of figure 6.6 using potential function V_2 equation (5.4).

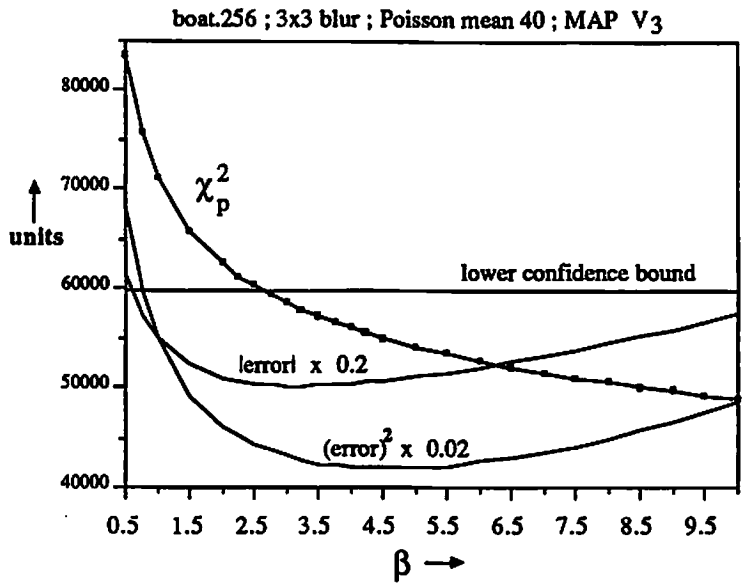

Figure 6.13
Plot of the L_1 and L_2 restoration error versus values of the Gibbs prior parameter in MAP restorations of figure 6.6 using potential function V_3 equation (5.5).

we would expect restorations of the moons surface to benefit from a greater influence from the Gibbs prior. Comparing the L_1 and L_2 restoration error for a wide range of β values as shown in these two graphs we see this is the case. A good β value for the restoration of the moons surface would be about 100. The statistical selection method results in a β of 125. However, a β value of 100 is far too small for the restoration of the image of the boats as the graph in figure 6.11 indicates and results in a high L_1 and L_2 restoration error. A β value in the range of 600 is much better. The statistical method of selecting β chose β equal to 500. A similar effect can be seen in MAP restorations of the two images using Gibbs priors with the other two potential functions in equations (5.4) and (5.5). Compare the two graphs in figures 6.6 and 6.12 and compare the two graphs in figures 6.7 and 6.13 for a further look at how the content of an image can affect whether a

particular β value is useful and at how the statistical method of selecting the β parameter can track this effect.

6.8 A Negative Feedback Algorithm

We have formulated the restoration and reconstruction problems to specify both the β value and the estimated image $\hat{\lambda}$. This approach relies upon a statistic $s[\lambda;y]$ satisfying two criterion and upon an appropriate confidence interval $[\gamma_1,\gamma_2]$ for that statistic. We require $s[\hat{\lambda}_0;y] > \gamma_2$ and $s[\hat{\lambda}_\infty;y] < \gamma_1$ where $\hat{\lambda}_\beta$ is the estimate after convergence of a MAP algorithm using a Gibbs prior with prior parameter β. Suitable statistics for the Poisson data model and the linear Gaussian data model have been specified. For the sake of deriving an effective algorithm which simultaneously selects the proper value for β such that $s[\hat{\lambda}_\beta;y]$ ε confidence interval we will cast the reconstruction process into a control systems framework.

First we make the assumption that $s[\hat{\lambda}_\beta;y]$ versus β is, or is well modeled as, a continuous function of β for the region of β in which the statistic is equal to the lower confidence bound. This is to insure that a solution as posed exists and that a solution is achievable by a deterministic approach. For a given Gibbs prior, the location of the global maximum of the posterior density function, as well as the locations of local maxima, need not be continuous functions of β. However, the mapping from the estimate corresponding to a local maxima for a given β value, to the statistic value $s[\lambda^k_\beta;y]$ has such a dramatic smoothing effect. From the previous section, we have $s[\hat{\lambda}_0;y] > \gamma_2$ and $s[\hat{\lambda}_\infty;y] < \gamma_1$. We therefore postulate a plot of $s[\hat{\lambda}_\beta;y]$ versus β of some form like figure 6.14.

Lets examine a deterministic MAP algorithm more closely. Given a set of data y, a Gibbs prior, a selected prior parameter β_1, and an initial estimate $\lambda^0_{\beta_1}$, the

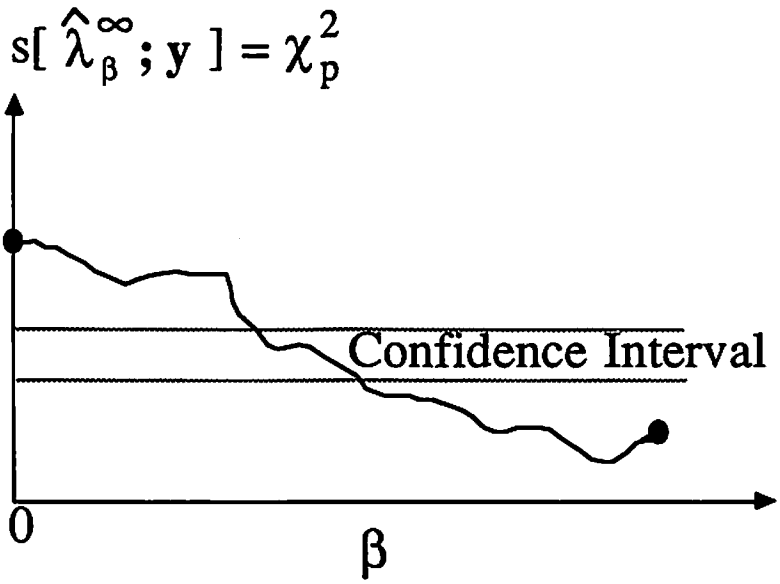

Figure 6.14
Knowing two points on the graph, $\beta = 0$ and $\beta = +\infty$, $s[\hat{\lambda}_\beta;y]$ versus β is modeled as being continuous in the region $s[\hat{\lambda}_\beta;y] =$ the lower confidence bound.

optimization process proceeds in a deterministic manner and converges to some local optimum $\hat{\lambda}_{\beta_1}$. We could evaluate the chosen statistic at that point $s[\hat{\lambda}_{\beta_1};y]$ and check whether it lies within a chosen confidence interval. Given an alternate β value, β_2, the MAP algorithm would converge to $\hat{\lambda}_{\beta_2}$ and we could evaluate $s[\hat{\lambda}_{\beta_2};y]$. Given $\hat{\lambda}_\beta$ for a number of β values, we could form a plot of $s[\hat{\lambda}_\beta;y]$. This is what we have done in figures 6.5, 6.6, 6.7, 6.11, 6.12, and 6.13. We would then choose the largest β value for which $s[\hat{\lambda}_\beta;y]$ lay within the chosen confidence interval. The MAP estimate $\hat{\lambda}$ corresponding to that β value would be the estimate of the image.

To cast the reconstruction problem in a control system framework, we first view β as the input to a system with output $s[\hat{\lambda}_\beta;y]$ as in figure 6.15. When a β

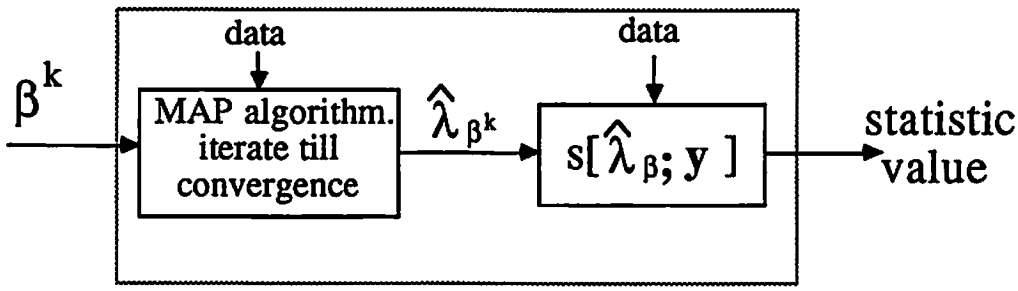

Figure 6.15
A system representation of the MAP restoration process with input β and output statistic value $s[\hat{\lambda}_\beta;y]$.

value is input, the plant performs the iterations of the MAP restoration algorithm until convergence to $\hat{\lambda}_\beta$ and then computes the statistic $s[\hat{\lambda}_\beta;y]$. If we knew the correct β value to ensure $s[\hat{\lambda}_\beta;y] = \gamma_1$, we would input β and let the deterministic MAP algorithm converge to the $\hat{\lambda}_\beta$ for which this is satisfied. One could imagine a process in which a β value is input, the MAP algorithm is allowed to converge to $\hat{\lambda}_\beta$, and $s[\hat{\lambda}_\beta;y]$ is evaluated. If $s[\hat{\lambda}_\beta;y] < \gamma_1$ we decrease β and repeat the process. If $s[\hat{\lambda}_\beta;y] > \gamma_2$ we increase β and repeat the process until the condition is satisfied. However, an approach such as this would be computationally intractable.

In order to be more efficient, we will let the input β vary at each iteration of the optimization process, i.e. input β becomes β^k at the k^{th} iteration, so that the output of the system becomes $s[\hat{\lambda}^k_\beta;y]$ at the k^{th} iteration. The system then takes the form of figure 6.16 in which the plant performs only one iteration of the MAP

algorithm and then evaluates the statistic $s[\lambda_{\beta^k};y]$

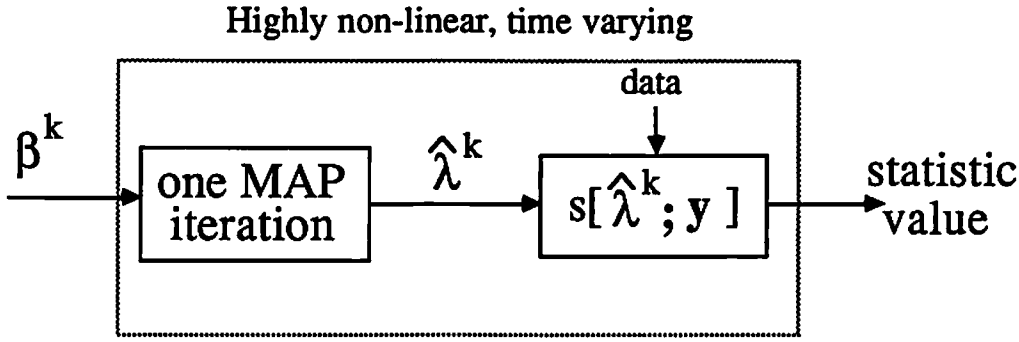

Figure 6.16
A system representation of the MAP restoration process in which β is allowed to change at each iteration of the MAP algorithm. The plant performs one iteration of the MAP algorithm then computes the statistic $s[\lambda_{\beta^k};y]$.

This system model is both non-linear and time-varying. To date, there are no adaptive control methods for non-linear time-varying systems which can be shown to be stable. Instead, we take a model reference control approach by posing a model for the convergence behavior of the statistic and specifying an algorithm for adjusting β^k when the system is not performing within bounds set with respect to the reference model (figure 6.17).

For any given β^k value, an iteration of a deterministic MAP algorithm generates an updated estimate λ_β^k at which the statistic $s[\lambda_\beta^k;y]$ is evaluated. Since $\lambda_\beta^k \rightarrow \hat{\lambda}_\beta$, $s[\lambda_\beta^k;y] \rightarrow s[\hat{\lambda}_\beta;y]$. For the reference model, we pose the following convergence behavior of the statistic value :

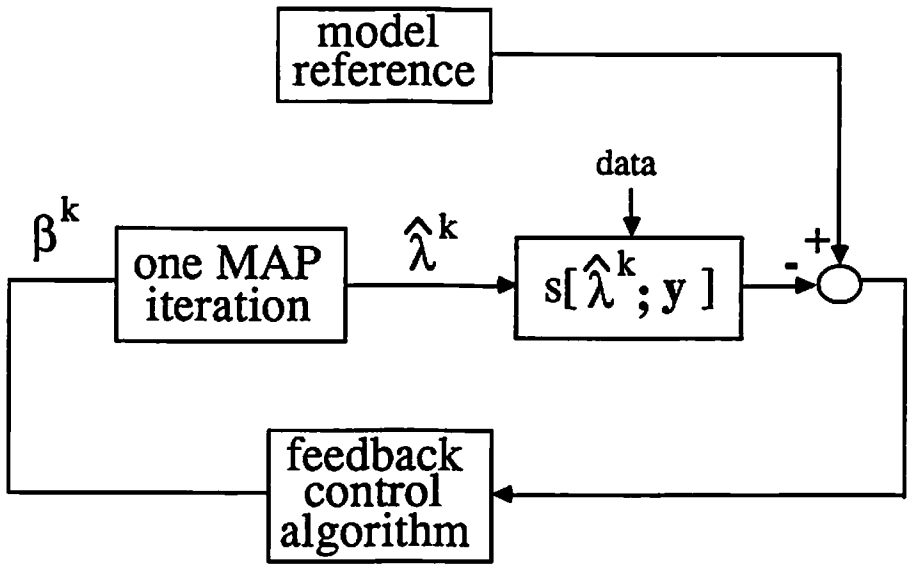

Figure 6.17
Model reference control of the plant in figure 6.16.

Given any initial estimate λ^0 , the values of the statistic resulting from iterations of the deterministic MAP algorithm has the form

$$s[\lambda_\beta^k;y] = \alpha^k s[\lambda_\beta^{k-1};y] + (1-\alpha^k)s[\lambda_\beta^\infty;y] \tag{6.21}$$

i.e. the value of the statistic decays with respect to iteration towards $s[\lambda_\beta^\infty;y]$. We allow the model reference system a range of acceptable rates α^k for convergence of the statistic towards the lower confidence bound γ_1, where

$$\alpha^k = \frac{s[\lambda_\beta^k,y] - \gamma_1}{s[\lambda_\beta^{k-1},y] - \gamma_1} \tag{6.22}$$

Such a range $[\alpha_1,\alpha_2]$ may be [0.3,0.7]. The feedback control block if figure 6.14 adjusts β^{k-1} if the convergence rate is not in the acceptable range.

The stucture then becomes one of a non-linear control problem. The control signal at the k^{th} iteration β^k is input into the first non-linear block of figure 6.17

which performs one iteration of a deterministic MAP algorithm and outputs λ^k. The second non-linear block receives λ^k and outputs the statistic $s[\lambda_{\beta^k}^k, y]$.

The control block computes the convergence rate α^k as in equation (6.22). If α^k falls within the $[\alpha_1, \alpha_2]$, $s[\lambda_{\beta^k}^k, y]$ is assumed to be converging towards the correct value $\hat{v}^k = \gamma_1$. If $\alpha^k < \alpha_1$, it is assumed that $s[\lambda_{\beta^k}^k, y]$ is converging towards a value $\hat{v}^k$ below its confidence interval, at rate $\alpha^k = \alpha_1$ so that

$$\hat{v}^k = \frac{s[\lambda_\beta^k; y] - \alpha_1 \, s[\lambda_\beta^{k-1}; y]}{(1-\alpha_1)}. \tag{6.23}$$

If $\alpha^k > \alpha_2$, it is assumed that $s[\lambda_{\beta^k}^k, y]$ is converging towards a value $\hat{v}_L$ above its confidence interval, at rate $\alpha^k = \alpha_2$ so that

$$\hat{v}^k = \frac{s[\lambda_\beta^k; y] - \alpha_2 \, s[\lambda_\beta^{k-1}; y]}{(1-\alpha_2)}. \tag{6.24}$$

Since β must remain strictly positive, we chose multiplicative feedback based on the error $e^k \gamma_1 - \hat{v}^k$.

$$\begin{aligned} &\text{if } \hat{v}^k = \gamma_1 \quad \beta^k = \beta^{k-1} \\ &\text{if } \hat{v}^k > \gamma_1 \quad \beta^k = \beta^{k-1}\{5.0 - 4.0\exp[+4.0\, e^k/\gamma_1]\} \\ &\text{if } \hat{v}^k < \gamma_1 \quad \beta^k = \beta^{k-1}\exp[-4.0\, e^k/\gamma_1)] \end{aligned} \tag{6.25}$$

At each iteration, β^k is updated by multiplying it by some number between 0 and 5.0 based upon acceptable convergence rates of $s[\lambda_\beta^k; y]$ towards the desired value γ_1 and a simple estimate of the value of the value $\hat{v}^k$ towards which $s[\lambda_\beta^k; y]$ is converging otherwise. This function is shown in figure 6.18.

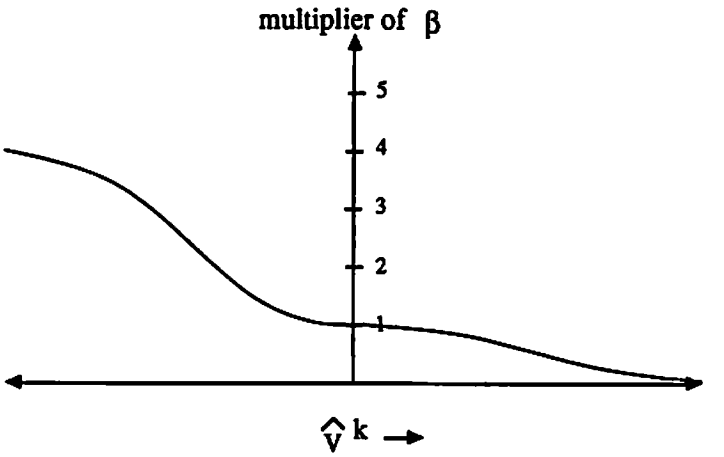

Figure 6.18
Multiplicative updating function for β^k based upon a simple estimate of the value to which $s[\lambda_\beta^k;y]$ is converging. Desired convergence value is γ_1.

6.9 Performance of the Algorithm and Conclusion

The following four graphs, figures (6.19)-(6.24), demonstrate the performance of the algorithm for achieving the solution posed in this chapter. The solution arrived at is as displayed in the examples shown in section 6.5. The simple regulator used in conjunction with the model reference control approach was effective in insuring convergence of both the stattistic and β to their desired values. As can be seen from the figures, MAP restoration using a Gibbs prior with the statistic based method of prior parameter selection can be implemented so that convergence is quite rapid. The added computation is minimal and the benefits of this approach can be judged from both the quality of the reconstruction in the previous section and from the L_1 and L_2 reconstruction error resulting from alternative values of β being used in the MAP algorithm.

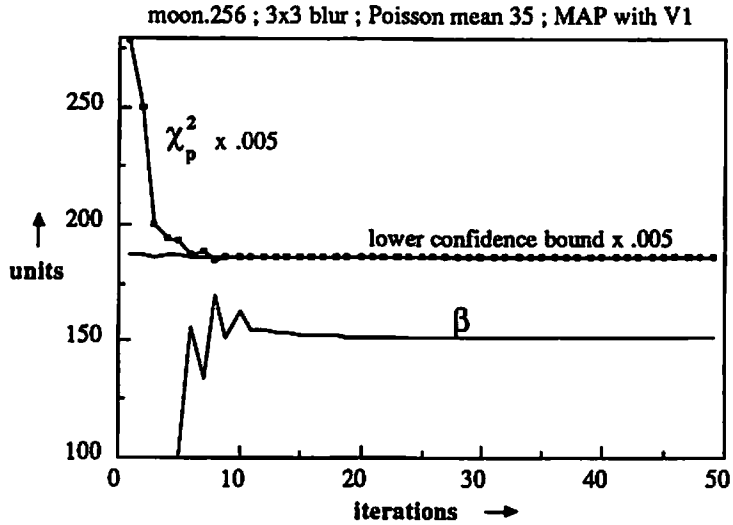

Figure 6.19
Performance of the regulator causing convergence of β^k to the value for which the statistic equals the lower bound in figure 6.5. The restoration is the bottom of figure 6.3.

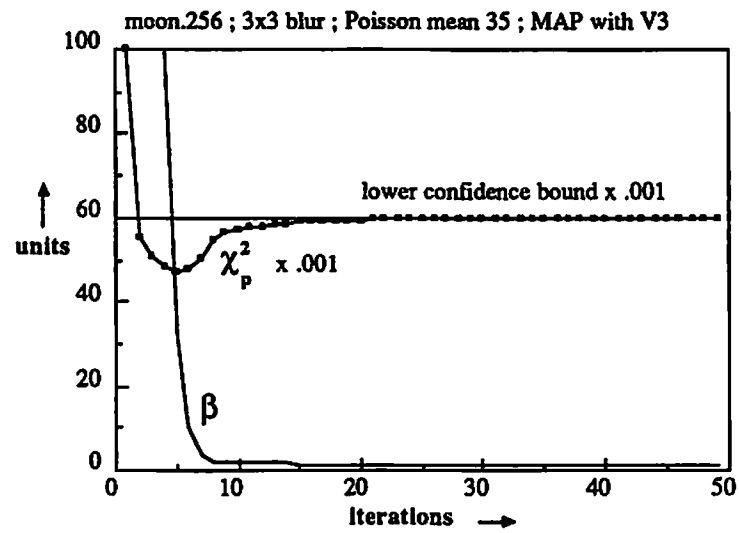

Figure 6.20
Performance of the regulator causing convergence of β^k to the value for which the statistic equals the lower bound in figure 6.7. The restoration is the bottom of figure 6.4.

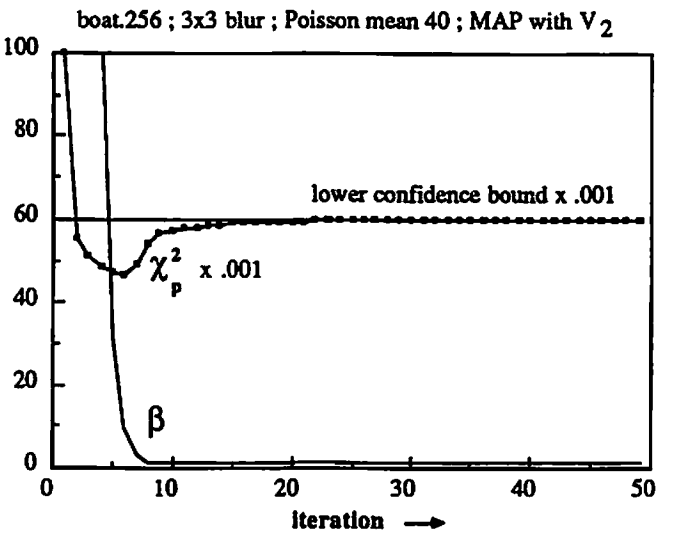

Figure 6.21
Performance of the regulator causing convergence of β^k to the value for which the statistic equals the lower bound in figure 6.12. The restoration is the top of figure 6.10.

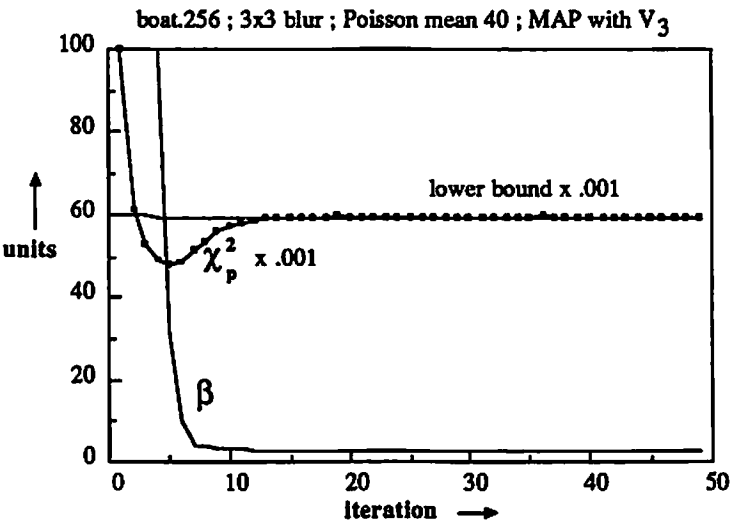

Figure 6.22
Performance of the regulator causing convergence of β^k to the value for which the statistic equals the lower bound in figure 6.11. The resulting reconstruction is as appears in the bottom of figure 6.13.

Chapter 7

Applications and Implementations in Emission Tomography

7.1 Introduction

This chapter deals with modeling, fast implementations, and increased convergence speeds for emission tomographic reconstruction. A method of reducing the computations required to implement iterative reconstruction algorithms for mechanically collimated emission imaging systems is presented in section 7.2. In section 7.3, a computationally frugal method of incoporating attenuation directly into the ML reconstruction approach is presented. In section 7.4, a stochastic model for a prototype imaging system is developed. This system has a spatially varying point source response and generates enormous amounts of data, 20.7 x 10^6 values, to reconstruct a 3-D space of 48^3 image pixels. A factorization of the system matrix is derived to speed implementation of iterative algorithms by several orders of magnitude. A modification of the standard EM ML algorithm is introduced to increase the convergence speed by a factor of five.

7.2 Use of Intermediate Polar Pixel Representation

It has been argued that for statistical reasons, only pixel representations with equal pixel areas should be used [41] in emission tomography. The problem lies in the size disparity between the polar pixels. Where the center pixel is omitted and there are n_r pixels radially from the center of the grid to the edge, the outermost pixels are $(2n_r-1)$ times larger than the innermost pixels. This would suggest either the innermost source pixels are below the resolution of the system or the outermost pixels are too large for the resolution of the system. If the isotope

distributions are represented with polar pixels, the reconstructions have a SNR which varies spatially from the center of the reconstruction to the perimeter.

However, using rectangular-to-polar transformations, the actual estimation can be carried out in terms of rectangular source pixels, while an intermediate polar source representation is used to perform forward/back projections [34]. Numerical agreement between forward/back projections with polar pixels and those with rectangular pixels can be achieved to the desired accuracy by simply increasing the number of polar pixels. This is aided by the fact that typical emission source profiles are non-zero only within some center region of the reconstruction space. Within this center region, the size of the polar pixels is smaller than that of the rectangular pixels. The intermediate polar pixels serve simply as a book keeping device during forward/back projection.

7.2.1 Pixelated Images

A rectangular or polar source representation models continuous distributions under the assumption of homogeneity within each pixel. The value assigned each pixel signifies the integral of the distribution over the volume of the pixel. It is therefore reasonable to form the value of a polar pixel from a rectangular-polar transformation as the integral of the rectangular pixel representation over the volume of the polar pixel. Let a rectangular pixel be of unit volume. Then, the value of a polar pixel is equal to the sum of the products of rectangular values with the fraction of their volume overlapping the polar pixel (figure 7.1), a pixel-overlap interpolation method. In matrix-vector notation, this can be represented as follows. Let λ_r and λ_p be vectors containing the lexicographically ordered rectangular and polar pixels respectively. Then a rectangular-to-polar transformation is represented

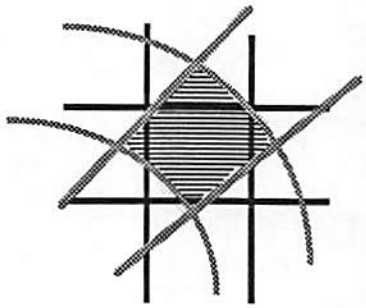

Figure 7.1
Overlap between a polar pixel and rectangular pixels.

by $\boldsymbol{\lambda}_p = \mathbf{T}_{rp}\boldsymbol{\lambda}_r$ where $\mathbf{T}_{rp}$ is the transformation matrix containing the overlapping volumes between the two pixel representations. Alternatively, a polar-to-rectangular transformation is represented by $\boldsymbol{\lambda}_r = \mathbf{T}_{rp}{}^t\mathbf{V}\boldsymbol{\lambda}_p$ where $\mathbf{V}$ is a diagonal matrix with

$$\mathbf{V}(i,i) = \frac{\text{volume of a rectangular pixel} = 1}{\text{volume of polar pixel } i} \tag{7.1}$$

Since no polar pixel intersects more than a few rectangular pixels, $\mathbf{T}_{rp}$ is very sparse. The transformations can be implemented with a single look-up table, and the required number of multiplies is only several times the number of polar pixels.

7.2.2 Reducing Computations

When a polar source representation is used, the mapping from the mean the source image to the mean of the data has a rotational invariance with respect to camera viewing angle (figure 7.2). This gives a redundancy to the forward projection matrix $\mathbf{P}$ which can be arranged to have a block-circulant structure. Let the polar representation consists of n_a pixels around 360° and n_r pixels radially from center, such that there is no center pixel. Let the data consist of $n = n_a/k$ views over 360°; k, n integers ($k = 1,2$), with m data pixels per row per view. For a single transverse slice, if we form a source vector by concatenating the radial columns of

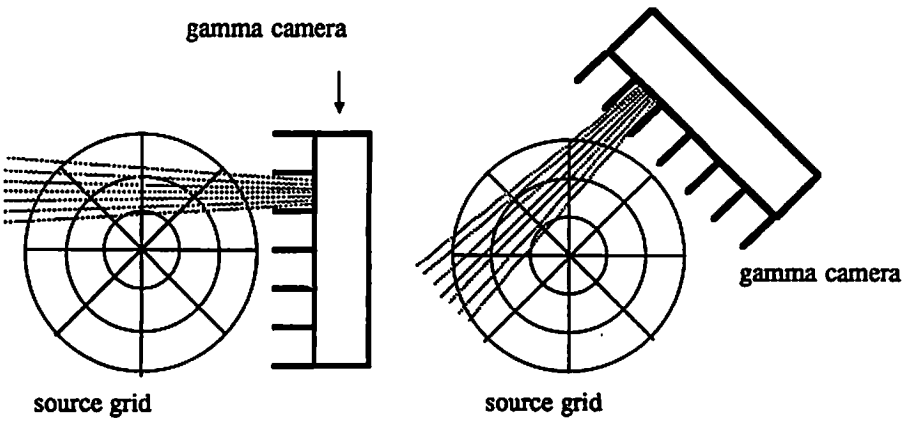

Figure 7.2
Mapping from the mean of the source image to the mean of the data pixel as seen at two camera viewing angles.

the source and form a data vector by concatenating the row columns of the data, the resulting $(nm)\times(n_a n_r)$ system matrix has a block-circulant structure with rectangular $(m)\times(kn_r)$ blocks. The matrix factorization for square block-circulant matrices [24] easily extends to this rectangular form. Therefore, $\mathbf{P}=\mathbf{W}_d^{-1}\mathbf{D}\mathbf{W}_s$ where

$$\mathbf{P}=\begin{bmatrix} [P_1][P_2]..[P_n] \\ [P_n].....\\ ..... \\ ... ... \\ [P_2]..[P_n][P_1] \end{bmatrix} \qquad \mathbf{D}=\begin{bmatrix} [D_1] & & [0] \\ & .. & \\ & [D_i] & \\ & & .. \\ [0] & & [D_n] \end{bmatrix} \tag{7.2}$$

and $\mathbf{D}$ is a $(nm)\times(n_a n_r)$ complex block-diagonal matrix whose rectangular blocks $\mathbf{D}_i$ are $(m)\times(kn_r)$. $\mathbf{W}_d$ and $\mathbf{W}_s$ are the block DFT matrices formed as the Kronecker matrix product

$$\mathbf{W}_d = \mathbf{W}_n \otimes \mathbf{I}_m \quad \text{and} \quad \mathbf{W}_s = \mathbf{W}_n \otimes \mathbf{I}_{n_r} \tag{7.3}$$

where $\mathbf{I}_m$ and $\mathbf{I}_{n_r}$ are $(m)\times(m)$ and $(n_r)\times(n_r)$ identity matrices respectively, and $\mathbf{W}_n(p,q) = \frac{1}{m}\exp\{\frac{-j2\pi pq}{n}\}$. The first block row can be pre-computed, and used to pre-compute and store the diagonalized matrix. Therefore, close attention can be paid to modeling and computing these probabilities accurately, increasing the integrity of the reconstruction accordingly. Then, the rectangular block-circulant structure of the probability matrix enables the use of FFT's in computing matrix-vector multiplications; i.e.

$$\mathbf{Px} = \mathbf{W}_d^{-1}(\mathbf{D}(\mathbf{W}_s\mathbf{x})) \quad \text{and} \quad \mathbf{P}^t\mathbf{y} = \mathbf{W}_s^{-1}(\mathbf{D}^*(\mathbf{W}_d\mathbf{y})) \tag{7.4}$$

This FFT based stucture is ideally suited for parallel processing.

An intermediate polar representation can be used to reduce the computational demands of many iterative reconstruction algorithms. For SPECT, the EM algorithm [64] is the most popular iterative ML algorithm at this time. A single iteration of the EM algorithm can be viewed as requiring four steps. Let $\boldsymbol{\lambda}_r$ be the vector of lexicographically ordered rectangular source pixels and $\mathbf{P}_r$ the projection matrix for those rectangular pixels. Then using the rectangular-to-polar transformation $\mathbf{T}_{rp}$ we can form the rectangular source estimate $\boldsymbol{\lambda}_r{}^{k+1}$ at the $k+1^{\text{th}}$ iteration from the rectangular source estimate $\boldsymbol{\lambda}_r{}^{k}$ at the k^{th} iteration by the following:

(1) forward project the estimated source means:

$$\mathbf{P}_r\boldsymbol{\lambda}_r = \mathbf{W}_d^{-1}\mathbf{D}\mathbf{W}_s\mathbf{T}_{rp}\boldsymbol{\lambda}_r^k \;;$$

(2) form the normalized data vector: $\hat{\mathbf{y}}_i = \dfrac{\mathbf{y}_i}{[\mathbf{P}_r\boldsymbol{\lambda}_r]_i}$;

(3) back-project the normalized data vector:

$$\psi = \mathbf{T}_{rp}{}^{t}\mathbf{V}\mathbf{W}_{s}^{-1}\mathbf{D}^{t}\mathbf{W}_{d}\hat{y}$$

(4) update the source estimate:

$$\lambda_{r}^{k+1}(i) = \frac{\lambda^{k}(i)\cdot\psi(i)}{[\mathbf{T}_{rp}{}^{t}\mathbf{V}\mathbf{P}^{t}\vec{1}]_{(i)}} = \frac{\lambda^{k}(i)\cdot\psi(i)}{[\mathbf{T}_{rp}{}^{t}\mathbf{V}\mathbf{W}_{s}^{-1}\mathbf{D}^{t}\mathbf{W}_{d}\vec{1}]_{(i)}}$$

where $\vec{1}$ is the $(nm)\times 1$ vector of 1's . Where an N element FFT can be computed in $N\cdot\log_2(N)$ operations, forward/back projection requires on the order of $n_a n_r[2\log_2(n_a)+mk]$ complex multiplies plus a rectangular-polar transformation. The total number of required operations, which is independent of the complexity of the forward projection matrix used, is slightly less that required for forward mapping of rectangular pixels using a forward projection in which every rectangular pixel contributes to only one camera pixel at every angle of view.

A reconstruction based on the intermediate use of polar pixels is shown in figure 7.3. The study is of radiolabeled isoamphetamine, a radiopharmaceutical which localizes in regions of brain activity. No artifacts result from using an intermediate polar pixel representation for the 3-D source.

7.3 Attenuation Compensation For SPECT Imaging Systems

No presentation on emission tomography would be complete without a discussion of attenuation effects. Attenuation of gamma rays by tissues within the body has for some time been recognized as one of the major factors affecting the data and thus the reconstruction [2],[12]. Attenuation of gamma rays takes place due to two processes, scattering and absorption. Scattering refers to gamma rays whose initial direction has been altered by a Compton scattering event. These gamma rays loose some of their initial energy in such an event. Since the amount of light

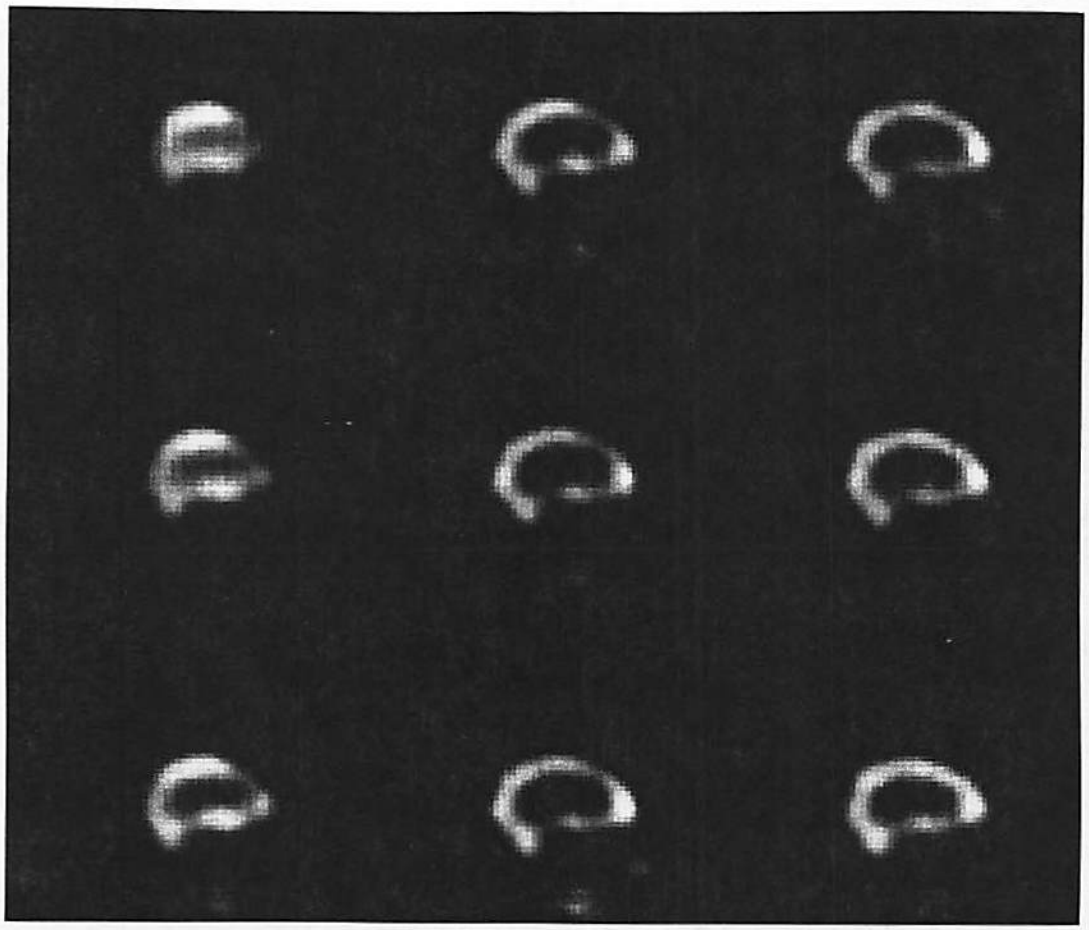

Figure 7.3
9 planes of a ML reconstruction of a head study using an intermediate polar pixel representation.

generated in a scintillation crystal by an impinging gamma ray is proportional to the energy of the gamma ray, to a large degree Compton scattered gamma rays can be omitted from the data by thresholding at the output of the photomultipliers. Absorption occurs when a gamma ray interacts with the lattice of the material

through which it is passing, transferring all its energy to the lattice in the form of heat. Both of these processes can reduce the number of gamma rays which would be counted at a gamma camera pixel were attenuation neglible. The probability that a gamma ray will be attenuated by either process as it passes through a material is dependent upon the material through which it is passing, the distance of its path through the material, and upon the energy of the gamma ray. For body tissues and widely used isotopes, these values are available from a body of experimental measurements, for instance the Radiological Health Handbook (U.S. Department of HEW).

Let μ_1 be the attenuation coefficient in units of cm^{-1} for a soft tissue and μ_2 be the coefficient for bone. Then for a gamma ray passing first through 3 cm of soft tissue followed by 2 cm of bone, we will express the probability that the gamma ray will not be attenuated as

$$\text{Prob(not attenuated)} = e^{-3\mu_1}e^{-2\mu_2} = e^{-(3\mu_1 + 2\mu_2)} \tag{7.5}$$

We will assume the attenuation coefficient for air is approximately zero.

7.3.1 The System Matrix with Attenuation

During the period of data aquisition by an emission imaging system, the number of emissions from the source pixels are well modeled as independent Poisson variables with mean vector λ. Let vector $\mathbf{y}^k$ contain the 2-D data image collected by the gamma camera at viewing angle k. Vector $\mathbf{y}^k$ are Poisson with means $\mathbf{P}^k\lambda$ where $\mathbf{P}_{ij}^k$ is the probability that a γ-ray from the j^{th} source pixel will generate a count in the i^{th} data pixel at viewing angle k. Matrix $\mathbf{P}$ is tyically called the system matrix. If attenuation is ignored, matrix $\mathbf{P}$ can be computed from the layout of the system; type of collimator, number of viewing angles, number of gamma

camera pixels, etc.. If we are to include attenuation effects in **P**, each probability $\mathbf{P}_{ij}$ can be broken into a product of two values, $\mathbf{P}_{ji} = \mathbf{P}_{ij}^{s}\,\mathbf{P}_{ij}^{a}$. $\mathbf{P}_{ij}^{s}$ can be calculated from knowledge of the system. However, $\mathbf{P}_{ij}^{a}$ represents the attenuation of emissions from source element i as they pass through the body tissues towards the gamma camera. Therefore it is dependent upon the attentuation coefficients throughout the 3-D source volume. The attenuation coefficients for a particular patient are obtainable through transmission scans or estimation. If the attenuation coefficients are known or estimated, what is the computational cost of including this information directly in the reconstruction instead of as an afterthought by a post-processing method.

The effect of directly including the attenuation probabilities in iterative reconstruction algorithms for PET systems is not too severe, since all sources generating counts within a given coincidence tube experience the same probability of attenuation. The computational effect is proportional to the number of data coincidence tubes. However, an attenuation terms must be calculated for each of the coincidence tubes, which may number 50,000 per source plane. Some PET systems have the ability to acquire transmission scans through which the attenuation coefficients can be obtained.

For SPECT systems something more drastic occurs; all the non-zero elements of **P** become unique. Given the attenuation coefficients for all volume pixels in the reconstruction space, the probabilities $\mathbf{P}_{ij}^{a}$ can be calculated by summing attenuations along the ray path from the j^{th} source pixel to the i^{th} data pixel. Let M be the number of source pixels in each plane and K the number of camera viewing angles. The simplest system model assumes that each source pixel contributes to only one data pixel at every camera viewing angle. In this case we would need to follow

MxK ray paths and sum the attenuation coefficients along each path in every source plane. This is an extremely computationally intensive approach since M is at least 64^2 and K is typically 64. Symmetry can reduce the computation of the ray paths by something less than one fourth. For every iteration, then, we would need to compute 65,536 ray paths and form 262,144 sums of attenuation coefficients along ray paths for every plane of the 3-D source. Due to constraints on the amount of computation time considered reasonable in a medical setting, attenuation is only incorporated in reconstruction algorithms by appproximation methods.

Several inexact approaches have been devised to compensate for the attenuation effects in SPECT. These methods generally involve post-processing of the reconstructions and can yield strong improvements over assuming attenuation to be negligible. In the previous section we considered the use of intermediate polar pixel representations to reduce the computations required by iterative algorithms. By consideration of polar pixel representations for both the source and the attenuation profile we can formulate a fast approach to including attenuation effects directly into iterative maximum likelihood and least squares reconstruction algorithms.

7.3.2 Forward/Back Projection with Polar Pixel Representations

The system matrix **P** is a matrix of probabilities which relates the mean of the source pixels $\boldsymbol{\lambda}$ to the mean of the data values **y** through linear relationship $E\{\mathbf{y}\} = \mathbf{P}\boldsymbol{\lambda}$. Matrix **P** represents the spatially varying point source response function of the system. Depending on the reconstruction algorithm, linear transformation **P** is used to forward-project the source, back-project the scaled or filtered data, calculate error weighting coefficients, or all of the above. Storage, calculation, and

implementation of this linear transformation determines, sometimes entirely, the speed of the algorithm. When attenuation effects are assumed uniform, circular symmetric, ignored, or incorporated by post-processing approaches, the linear transformation $\mathbf{P}$ has a redundancy which can be exploited for its storage and implementation. In the previous section, 3.2, we saw that when polar pixels are used, $\mathbf{P}$ can be arranged to have a block-circulent structure. Forward and back-projections can be performed with dramatic computational efficiency by using FFT's to reduce $\mathbf{P}$ to a diagonal block strucure. However, if we try to directly incorporate a known set of attenuation coefficients which is neither uniform nor circular symmetric over the 3-D source volume, the strucure in $\mathbf{P}$ disappears. The method we will present in this section is a simple idea based on polar pixel representation of the attenuation coefficients.

Let vector $\mathbf{y}^k$ contain the data from the gamma camera at the k^{th} viwing angle. For a parallel-hole collimated SPECT system, a simple model for the contribution of the source pixels, vector $\boldsymbol{\lambda}$, to the data vector $\mathbf{y}_k$ is as shown in figure 7.4 where a source pixel is assumed to contribute to a data pixel only when some portion of the source pixel is within the viewing tube of the camera pixel. Every j^{th} pixel in the source volume has associated with it a mean $\boldsymbol{\lambda}_j$ for the isotope concentration and an attenuation coefficient a_j, both of which may be zero. At the first camera viewing angle, the rectangular source grid is lined up with the camera pixels. For a simple model, each source pixel therefore contributes to only one data value; i.e. where K is the total number of viewing angles, there are only K non-zero elements $\mathbf{P}_{ij}$ for each rectangular source pixel. Geometrically, a non-zero element is the probability that a γ-ray will head towards the data pixel located past the end of its row $\mathbf{P}_{ij}^s$, times the probability $\mathbf{P}_{ij}^a$ that it will pass unattenuated through

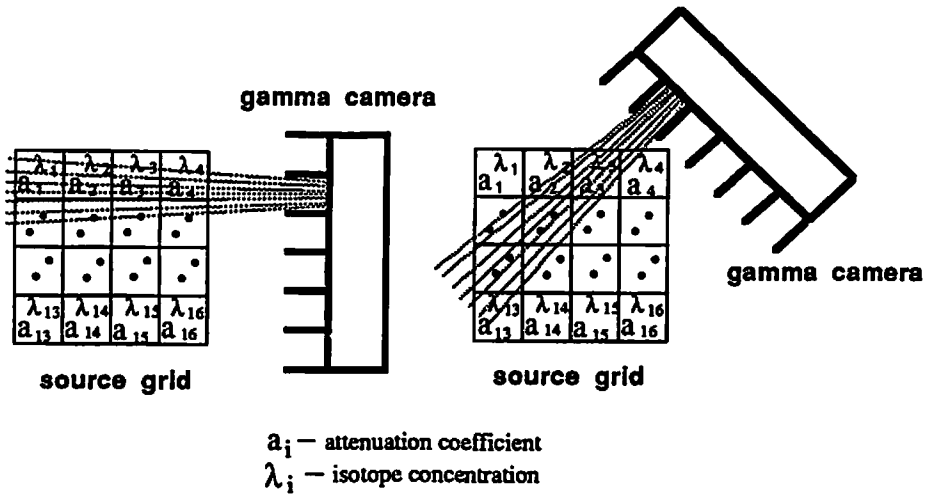

Figure 7.4
Geometry of the source and camera at the first two camera viewing angles using rectangular pixel representations of the isotope source and attenuation coefficients.

the pixels between it and the end of the row. At each viewing angle, since a source pixel contributes to only one camera pixel, we can store the non-zero probabilities $\mathbf{P}_{ij}^{s}$ in the form of an image, with the probability assigned each source pixel taking the location of the source pixel. The attenuation probabilities $\mathbf{P}_{ij}^{a}$ can be easily calculated by summing the attenuation coefficients between each rectangular pixel and the end of its row and using the form of equation (7.5). These can be multiplied to the probability image $\mathbf{P}_{ij}^{s}$. To forward project the mean of the source to the mean of the camera pixels we now only need to multiply the probability image times the estimate λ of the source and sum along the rows; i.e. $E\{y_j^k\} = \sum_{n=1}^{N} P_{jn}^{k} \lambda_{k,j}$. Attenuation has thus been included in the forward mapping of the source to the data space for the first camera viewing angle. We would like to maintain the simple structure of the first viewing angle for all viewing angles to avoid the complication of tracing ray paths and handling attenuation sums for every ray path.

However, proceeding to the second camera viewing angle, figure 7.5, we are faced with just this complication.

The problem is simplified by forming a new set of rotated rectangular source pixels, a new set of rotated attenuation coefficients and using the same strucure as that for the first viewing angle, figure 7.5. This is done quite simply.

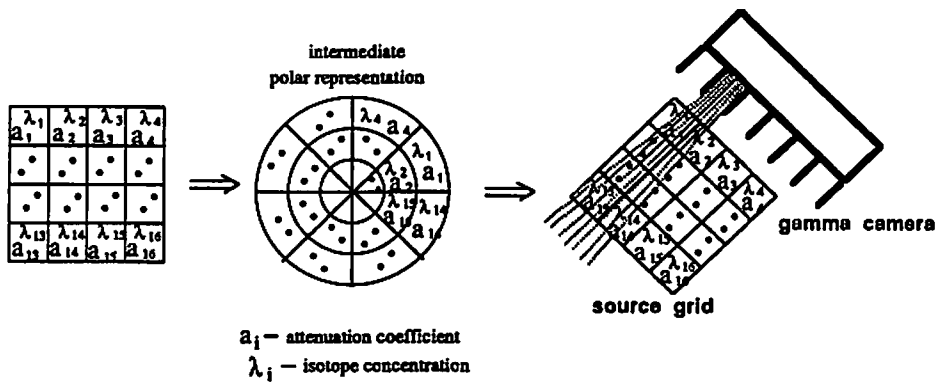

Figure 7.5
Geometry of source and camera for the second viewing angles using polar pixel representations for the source and attenuation coefficients.

Given the attenuation coefficients in the form of a rectangular image, the rectangular image is transformed to a polar pixel representation and stored. Given the current estimate of the source pixels λ, to forward project the source estimate to the mean of the data vector $\mathbf{y}^k$, image vector λ is transformed to polar pixels and stored. At each viewing angle, the polar pixel representations of both source λ and attenuation coefficients $\mathbf{a}$ are rotated so that a polar-to-rectangular transformation results in a rectangular grid oriented towards the gamma camera.

A back-projection, vector λ_{BP}, of scaled data values $\hat{\mathbf{y}}^k$ is formed in a similar way. At camera viewing angle k, a rectangular pixel image facing the direction of

the gamma camera and containing the non-zero elements of $\mathbf{P}_k$ is formed. Each scaled data value $\hat{y}_i^k$ is multiplied times every probability image value in its viewing row (the back-projection). This rectangular image is broken into polar pixels, rotated, and reassembled as rectangular pixels at zero viewing angle and summed with the images from the other camera viewing angles to complete the back-projection. For the simplified system model, inclusion of attenuation in froward/back projection operations becomes a computationally feasible process.

Again, the polar pixel representations only serve as an intermediate step to simplify and reduce the computations. When applied in an iterative reconstruction algorithm, the back-projections are always returned to rectangular pixel form to update the rectangular representation of the source.

7.3.3 The EM Algorithm with Direct Attenuation Compensation

Let vector $\mathbf{y}^k$ contain the 2-D data image aquired by a mechanically collimated gamma camera at the k^{th} viewing angle and let $\mathbf{P}^k$ be the probability matrix projecting the rotated rectangular source means to the mean of $\mathbf{y}^k$. Where λ_i^{n+1} is the value of the estimate of the i^{th} source pixel mean at the $n+1^{\text{th}}$ iteration, a single iteration of the EM algorithm for maximum likelihood estimation of $\boldsymbol{\lambda}$ from the Poisson data $\mathbf{y}$ can be represented as

$$\lambda_i^{k+1} = \frac{\lambda_i^k}{\left(\sum_{k=1}^{K}\sum_{m=1}^{M} \mathbf{P}_{mi}^k\right)} \sum_{k=1}^{K}\sum_{m=1}^{M} \frac{\mathbf{y}_m^k \mathbf{P}_{mi}^k}{\left(\sum_{l=1}^{N} \mathbf{P}_{ml}^k \lambda_l^k\right)} \qquad i=1,2,..N \qquad (7.6)$$

Implementation of a single iteration of the EM algorithm can be viewed as requiring five steps. For each of the K viewing angles:

1) forward project the estimated source means: $\mathbf{P}^k\boldsymbol{\lambda}$;

2) form the normalized data vectors $\hat{\mathbf{y}}^k$: $\hat{y}_i^k = \dfrac{y_i^k}{[\mathbf{P}^k\boldsymbol{\lambda}]_i}$;

3) back-project the normalized data vectors: $(\mathbf{P}^k)^t\,\hat{\mathbf{y}}^k$),

4) sum the back-projected images: $\psi_M = \sum_{k=1}^{K}(\mathbf{P}^k)^t\,\hat{y}_k$

5) update the source estimate: $\lambda_i^{k+1} = \lambda_i^k * \dfrac{\psi_M(i)}{(\sum_{k=1}^{K}\sum_{m=1}^{M}\mathbf{P}_k(m,i))}$.

For a known set of attenuation coefficients, the maximum likelihood estimate is obtained if attenuation is incorporated into the forward- and back- projections in steps 1) and 3). This can be done using the intermediate polar pixel representations as described above. Figure 7.6 compares this approach with results obtained using a popular post processing method of compensating for known attenuation. This simulation assumes perfect collimation over 64 viewing angles. In each row, from left to right, the images shown are the attenuation coefficients, the isotope source image, ML reconstruction without attenuation, ML reconstruction without attenuation post-processed by Changs method [12], ML reconstruction without attenuation post-processed by the Changs modified method [12], and ML without the attenuation coeficients directly incorporated into the algorithm. Although, Changs modified post-correction method performs quite well, ML reconstruction with the attenuation coefficients directly incorporated into the algorithm produced the best results.

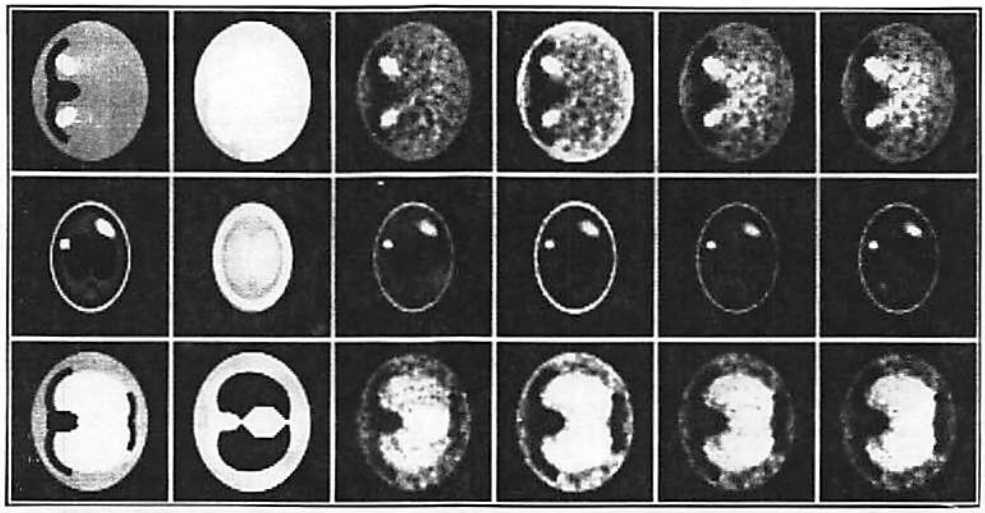

Figure 7.6
Reconstructions of 3 computer phantoms used in a simulation of a parallel hole collimated single photon emission system. In each of the three row, from left to right: 1) attenuation profile. 2) isotope concentration. 3) ML reconstruction without attenaution correction. 4) ML reconstruction followed by Changs post-correction method. 5) ML reconstruction followed by the Changs modified post-correction method. 6) ML reconstruction based on attenuation using the method presented in this section.

7.4 3-D MLE for a Prototype Emission Imaging System

In this section we present a 3-D maximum-likelihood reconstruction method for a prototype electronically collimated single photon emission system. The SPECT system under consideration in this paper uses a gamma camera fronted by an array of germanium (Ge) detectors in place of a conventional mechanical collimator. Only gamma rays which Compton scatter within a Ge detector and impinge upon the gamma camera are recorded. This performs an "electronic collimation". Such a system has a number of factors which present an interesting reconstruction challenge:

(1) The system generates extremely large amounts of data to be processed.

(2) The data are generated from highly complicated sampling geometries.

(3) The generally low count totals in most sample bins make Poisson fluctuations a significant factor.

Factor (1) imposes a formidable computational burden; (2) makes established analytical reconstruction techniques such as filtered-backprojection inapplicable; and (3) makes a maximum likelihood approach more desirable than a linear system or analytical approach [8],[10].

For the electronically collimated camera, we have formulated and implemented a direct 3-D ML reconstruction algorithm in a manner applicable to other systems. The system configuration we simulate generates over twenty million data values which are subsequently processed in the reconstruction. Through modeling considerations, quantization of several geometric parameters, and the intermediate use of a polar pixel representation when projecting the source estimate into the data space, the computation requirements of the EM maximum likelihood algorithm can be reduced by orders of magnitude for this highly complex system.

7.4.1 A Brief Description of the System

The physics of electronic collimation and the design of a system have been described in detail in [67],[68],[69] but we offer a brief discussion here. A simple configuration for an electronically collimated system is shown in figure 7.9 in which a gamma camera is fronted by an array of germanium (Ge) detectors instead of a conventional mechanical collimator.

We will first consider the data generating process from the perspective of a single point source and a single Ge detector. Consider a point source of an isotope

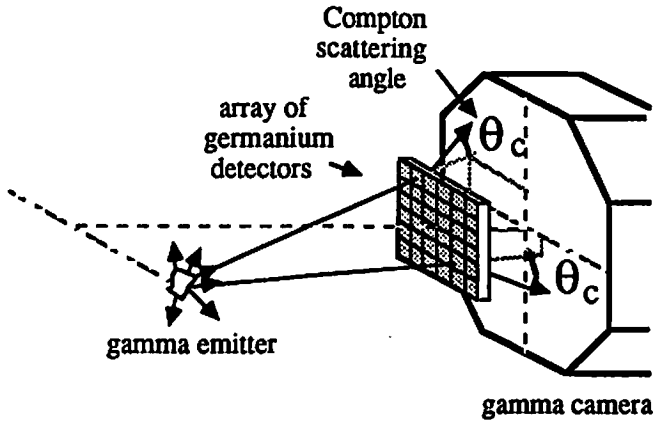

Figure 7.9
Diagram showing a configuration for an electronically collimated gamma camera. Pictured is a 6x6 array of germanium (Ge) detectors.

emitting gamma rays of a known energy as in figure 7.10. A gamma ray which Compton scatters within the Ge detector deposits energy proportional to the Compton angle of scatter θ_c. If, nearly simultaneously, the gamma camera detects a photon with approximately the correct remaining energy for the isotope, a count is recorded with three labels: the x,y coordinates of the photon detected by the gamma camera, the location of the Ge detector, and the Compton scattering angle measured by the Ge detector. However, due to the finite energy resolution of the Ge detector, the Compton scattering angle θ_c can be determined only to within some interval so that the recorded Compton angles are quantized values. From figure 7.10, the direction the photon takes after scattering is dependent upon the Compton scattering angle θ_c , upon arrival angles γ and ϕ and upon angle ψ which is random and uniformly distributed over 0 to 2π .

For each given point source location and Compton angle, there corresponds a 2-D probability image representing the probabilities of an emission from the source being detected at each camera pixel. Each 2-D probability image is equivalent to a

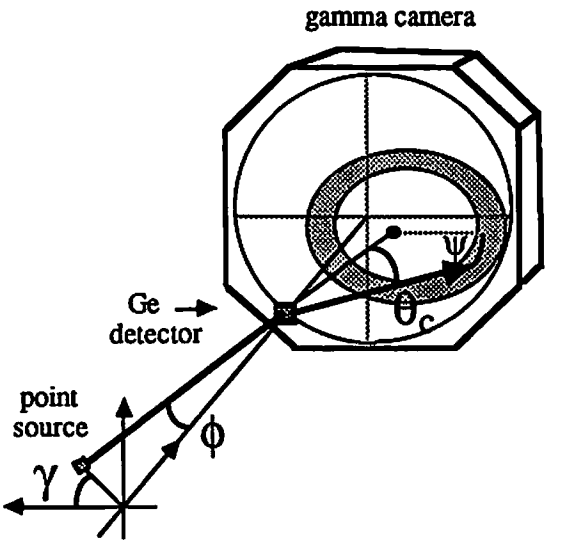

Figure 7.10
Point source, Ge detector, and the particular probability pattern for Compton scattering angle θ_c given the point source location. Shown are the arrival angles ϕ and γ and the exit angle ψ.

conic section of a hollow cone. The shape and location of the conic section is determined as shown in figure 7.10 by the arrival angles ϕ and γ and by the quantized Compton scattering angle θ_c being recorded. Angles ϕ and γ simply specify the relationship between the point source location vector and the direction vector normal to the Ge detector. The system matrix relating the source distribution to the data is defined by these probability images. For a given detector, all the counts corresponding to a given quantized Compton angle can be stored as a 2-D count image.

The electronically collimated system we simulate in this paper corresponds to a prototype system at the University of Southern California. It employs a 4x4 array of Ge detectors in front of an uncollimated gamma camera. This system is then used to view the source volume from multiple viewing angles over 360^o. At each

viewing angle, eight to sixteen 2-D photon count images may be stored for each Ge detector (one data image for each quantized scattering angle). For the purpose of simulation in this paper we assume 20 equispaced viewing angles over 360^o and detections from 8 Compton scattering angles stored per Ge detector. From the perspective of the reconstruction problem, we may equivalently consider a system of 20 gamma cameras, each with a 4x4 array of germanium detectors. Each Ge detector results in eight 90x90 photon count images as in figure 7.11 which shows the binning of counts according to Ge detector and Compton scattering angle. For the remainder of this paper, we shall refer to the system configuration in this manner. The reconstruction problem we have addressed can be stated as follows: given a set of 2-D photon count images totalling 20.7×10^6 data values, where each image corresponds to photons which Compton scattered from a particular detector and within a particular Compton angle interval, reconstruct the distribution of isotope in a 48^3 pixel volume.

7.4.2 System Matrix Factorizations

Let vector $\mathbf{y} \in R^M$ denote the recorded data, $\boldsymbol{\lambda} \in R^N$, the vector of pixel values representing the mean emission rates from the source pixels and $\mathbf{P} \in R^{M \times N}$, the matrix containing the probabilities of an emission from each pixel being detected in each data sample, i.e. $E[\mathbf{y}] = \mathbf{P}\boldsymbol{\lambda}$.

Iterative schemes for estimating $\boldsymbol{\lambda}$ from $\mathbf{y}$ require the computation of forward projections, $\mathbf{P}\boldsymbol{\lambda}$, and back projections, $\mathbf{P}^T\mathbf{y}$. In emission tomography, the system matrix $\mathbf{P}$ performs a projection from the 3D source volume to a set of 2D 'projection' images. For the electronically collimated system considered in this paper a 90x90 pixel image is collected for eight different Compton scatter angles for each

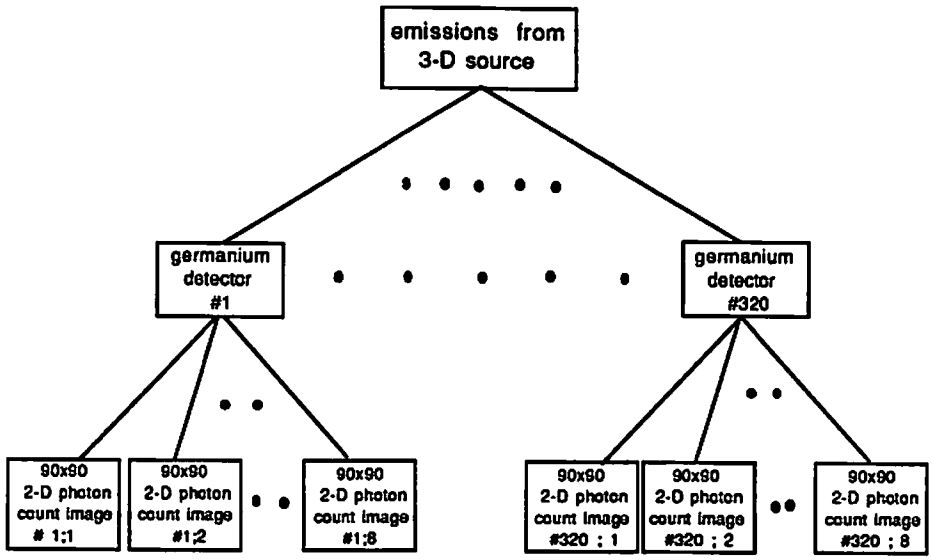

Figure 7.11
The binning of detected gamma ray emissions according to Ge detector they scattered in and Compton scattering angle.

of 320 Ge detector locations. The source volume has been divided in 48^3 pixels resulting in a system of size 20,736,000×110,592 . The storage of or multiplication by a matrix of this size would be unmanageable or impractical without some very efficient approach to the problem. In a previous publication [35], it has been shown how a matrix factorization can be exploited to reduce both storage and computation requirements for conventional mechanically collimated emission imaging systems. In the following section, we derive a matrix factorization based on the probabilistic mechanisms by which data is generated in the electronically collimated system.

7.4.2.1 Two Stage Factorization

In earlier papers [8],[9],[35] a two-stage linear reconstruction approach for the electronically collimated single photon emission computed tomography (SPECT) system was considered in which cone-beam projection images of the source volume were reconstructed, followed by estimation of the source image from the cone-beam projection images. This two stage approach may be interpreted in terms of a system matrix factorization as follows.

If the Ge detector were replaced with a simple pinhole collimator (a dense plate with a single central hole permitting the passage of γ-rays) then the mean of the data would represent an inverted cone-beam projection image of the mean of the source pixels. A cone-beam projection image [20] is simply a 2-D image whose pixel values represent volume integrals from a focal point through a 3-D intensity image. The focal point here would be the Ge detector location. The effect of replacing the pinhole collimator with the Ge detector is to map the inverted cone-beam projections through a transformation generating conical sections of a hollow cone from each element of the 2D projection image. This linear mapping results from the Compton scatter of photons passing through the Ge detector, where the Compton scatter angle determines the angle of the conic section (figure 7.9).

Let **B** be a matrix which projects the 3-D source image to a set of inverted 2-D cone-beam projection images and matrix **A** be a matrix which projects the set of 2-D inverted cone-beam projection images into the data space (i.e. through the transformation generating conical sections resulting from Compton scatter). Then, the system matrix **P** is equal to the product **AB**. In the following section, we show how the system matrix may be factored further by considering the probabilistic mechanism which relates the mean of the data to the mean of the source pixels.

7.4.2.2 A Probabilistic Factorization

As discussed in section 7.4.1, a single 2-D photon count image (figure 7.11) is a record of all photons which Compton scattered within a particular Ge detector at a particular quantized Compton angle and were recorded at the gamma camera. One 2-D count image per quantized Compton scattering angle is recorded for each of the 320 Ge detectors. Let us examine a single 2-D count photon count image representing photon counts of gamma rays which scattered within the a^{th} Ge detector at the b^{th} quantized Compton angle and were recorded by the gamma camera. Let $\mathbf{y}^{ab}$ be the vector of data representing this 2-D count image. The expected value $E\{\mathbf{y}^{ab}\}$ is equal to a particular linear transformation $\mathbf{P}^{ab}$ of the 3-D source. Given an emission from source pixel j, the element $\mathbf{P}_{ij}^{ab}$ of the matrix represents the probability that the emission will scatter from the a^{th} Ge detector at the b^{th} quantized Compton scattering angle and generate a count at camera pixel i. Each element $\mathbf{P}_{ij}^{ab}$ can be factored into three components: the probability of an emission from source pixel i striking the a^{th} Ge detector (p_i^a), the probability that an emission will Compton scatter at the b^{th} quantized Compton angle and escape from the detector (p_b) given it struck the detector, and the probability p_{ij}^{ab} of a Compton scattered gamma ray striking a particular camera pixel i given it originated at source pixel j and scattered at the b^{th} quantized Compton angle.

By dividing each element $\mathbf{P}_{ij}^{ab}$ into its three probability components, we can factor the linear transformation mapping the source image to the mean of the data vector into the matrix product form $\mathbf{P}^{ab} = p_b \mathbf{T}_{pr} \mathbf{H}^{ab} \mathbf{Q}^a \mathbf{D}^a$ where the p_b's are scalars, the p_i^a's are contained in the diagonal matrix $\mathbf{D}^a$, and the p_{ij}^{ab}'s are contained in the matrix product $\mathbf{T}_{pr} \mathbf{H}^{ab} \mathbf{Q}^a$ defined below. We will address each of these matrix components in turn and explain, in the paragraphs to follow, the

matrix product form which contains the p_{ij}^{ab}'s.

The direction of an emitted gamma ray from a point source is random and uniformly distributed over all directions in 3-D space. We assume that a post-attenuation correction method [12] will be applied to the reconstruction so that the probability that an emitted gamma ray will strike the Ge detector is initially independent of the attenuation profile. If the Ge detector is modeled as a sphere of known radius, then p_a is only dependent upon the distance from the point source to the detector. The probability component p_a is equal to the solid angle eclipsed by the spherical detector. Solution of this problem is straightforward and leads to $p_a = \frac{1}{2}[1 - \frac{\sqrt{d^2 - r^2}}{d}]$, where d is the distance from the point source to the Ge detector and r is the radius of the spherical model of the detector.

The second probability component p_b is the the probability a gamma ray which has struck a detector will Compton scatter at the b^{th} quantized angle. Given the Compton and photoelectric linear attenuation coefficients of the germanium detector, the incident photon energy for the isotope of interest, and the scattered photon energies for photons scattered at a given quantized Compton angle, the probability of a Compton scattering event within that quantized angle interval can be computed by integrating the Klein-Nishina distribution along the average path length over angles within the scattering angle interval [67]. An approximate normalized value for p_b can be formed from the ratio of the detected events from a particular quantized Compton scattering angle to the total detected counts:

$$p_b = p_C \frac{\text{\# of scattered events within a quantized Compton angle}}{\text{total number of recorded counts}}. \qquad (7.7)$$

Given that an emission from source pixel j struck the a^{th} germanium detector, scattered at the b^{th} quantized Compton angle and exited the detector, the third

probability component p_{ij}^{ab} is the probability of it striking a given pixel i of the gamma camera. This probability must be computed for each gamma camera pixel i. To compute this probability, the germanium detector is modeled as a point. Then, for a given quantized Compton scattering angle, all ray paths through the detector point which scatter from the detector within the given quantized Compton angle are assigned equal probability. For a given point source location, we can generate and store these third probability components using a computer program to trace a very large number of ray paths representing the uniform division of all possible ray paths from the point source through the given Ge detector and Compton scattering at the given quantized Compton angle onto the gamma camera pixels. For each pair (point source location , Compton scattering angle) these probability components can be stored as an image defined on the face of the gamma camera. For a fixed quantized Compton scattering angle, a typical set of probability images produced from a set of point sources with $\gamma = 0$ and ϕ increasing from 0 to $\frac{\pi}{3}$ as in figure 7.12a is shown in Figure 7.12b. These probability images are simply conic sections of a hollow cone. It would be impractical to compute and store these probabilities for each source pixel location, each Ge detector, and each quantized Compton angle, so that our handling of these probability components becomes central to a feasible implementation of a 3-D reconstruction algorithm.

A source pixel, detector, and the particular probability pattern for a given quantized Compton scattering angle has been shown in figure 7.10. By modeling the detector as a point source, the probability pattern is not dependent upon the distance from the source pixel to the detector, but only upon the angles ϕ and γ. For all source pixels, we quantize γ into 128 levels ranging from 0 to $\frac{127\pi}{64}$ and ϕ into 50 levels from 0 to the maximum arrival angle that can be generated from the 3-D

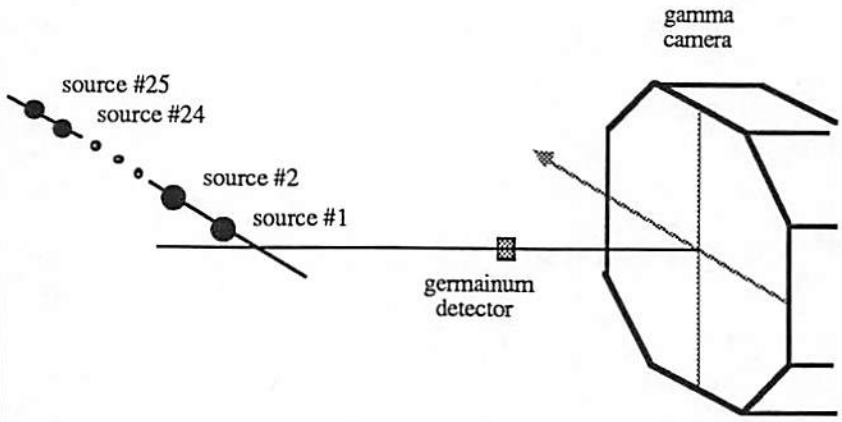

Figure 7.12a
Location of 25 point sources corresponding to the 25 probability point spread functions in figure 7.12b.

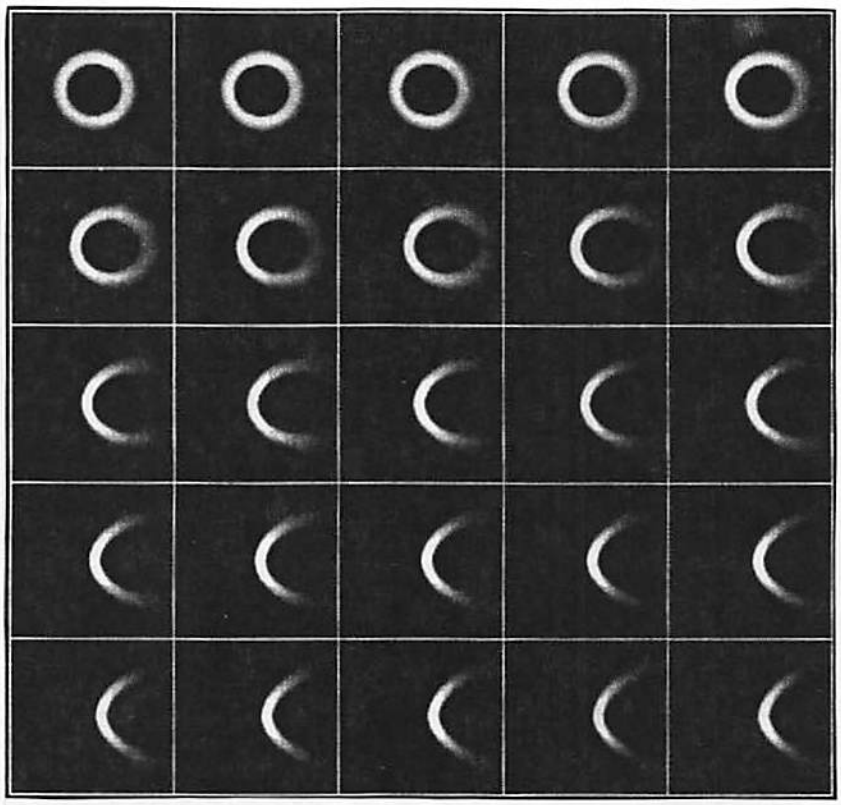

Figure 7.12b
25 PSF's associated with the 25 source locations in figure 7.12a. Begin with the top left hand image and proceed across the row.

source region. For the a^{th} Ge detector and the b^{th} quantized Compton scattering angle, this binning of source pixels according to quantized arrival angles ϕ and γ is a linear transformation represented here by $\mathbf{Q}^a$. This matrix simply effects a cone-beam projection of the 3-D source onto a polar raster where the a^{th} Ge detector is the focal point.

However, there is a more important computational benefit which can now be derived. The probability pattern from a point source with a ray path defined by γ_0 and ϕ_0 is simply a rotated version of that from a point source with ray path ϕ_0 and $\gamma_0 + \frac{2\pi}{128}$ (figure 7.13). Therefore, if we represent the probability image at the face of the gamma camera in terms of polar pixels with the number of polar pixels around the circle equal to the number of discretized levels for ψ, then the probability patterns for all source locations at a given angle ϕ can be reproduced by rotating the pattern from a source located at angle $\gamma = 0$. This is demonstrated in figure 7.13a and b. Further, the matrix $\mathbf{H}^{ab}$ which contains these probability patterns can thus be arranged to have a block-circulant structure. This structure is achieved by forming the polar probability image as a vector by concatenating polar pixels with the same angular location into sub-vectors, then concatenating those sub-vectors into a final vector. The rectangular pixel probability images at the face of the gamma camera can be formed from the polar probability images by a simple polar-to-rectangular transformation $\mathbf{T}_{pr}$ as described in [34]. Let the matrix $\mathbf{T}_{pr}$ represent a mapping from a polar pixel image to a rectangular pixel image. Then, the rectangular image vector $\mathbf{r}$ is related to its polar representation $\mathbf{p}$ by $\mathbf{r} = \mathbf{T}_{pr}\mathbf{p}$. Using this approach, we need only store the probabilities p_{ij}^{ab}, for the 50 quantized values of γ for each of the 8 Compton scatter angles. The elements are identical for each Ge detector and each of the 256 quantized values of ϕ.

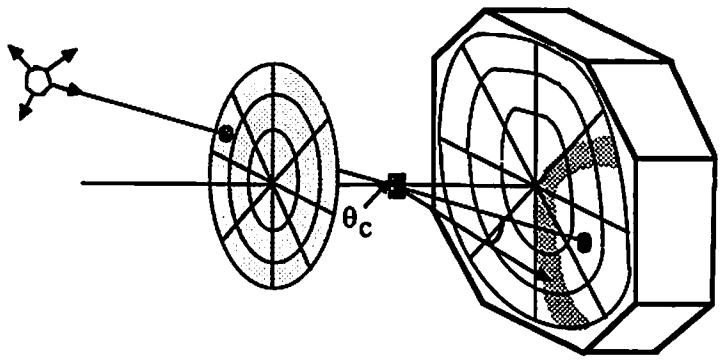

Figure 7.13a

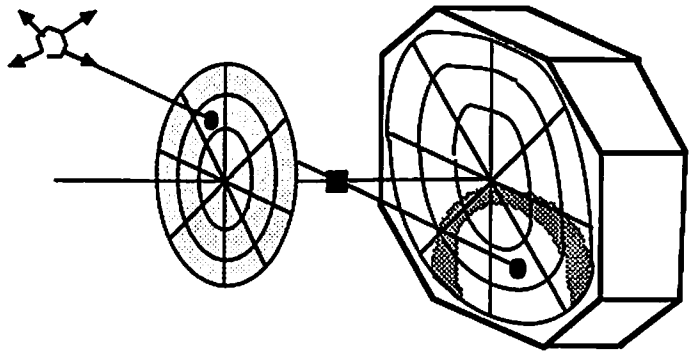

Figure 7.13b
This figure shows the rotational invariance of the probability image with respect to angle of arrival γ when the gamma camera is represented on a polar raster. The probability image on the polar raster at the face of the gamma camera in figure 7.13b is a rotated version of that in figure 7.13a.

In addition, to reducing the storage requirements for this probability, this block-circulant structure of $\mathbf{H}^{ab}$ can can be exploited to reduce the computational demands of forward/back projections as follows. The matrix factorization for square block-circulant matrices is shown in section 7.2. Using this we can factor $\mathbf{H}^{ab} = \mathbf{W}_d^{-1}\mathbf{R}^{ab}\mathbf{W}_s$ where $\mathbf{R}^{ab}$ is a complex block-diagonal matrix and $\mathbf{W}_d$ and $\mathbf{W}_s$ are the block DFT matrices as in equation (7.3). Multiplication by the block DFT matrices $\mathbf{W}_m^{-1}$ or $\mathbf{W}_n$ in this case is equivalent to taking the IDFT or DFT of the polar pixels at a fixed radius from the center of the polar image. This can be done very efficiently using established FFT routines.

Where λ is the vector containing the 3-D rectangular source representation, the rectangular estimate of the mean of a data set $E\{\mathbf{y}^{abc}\}$ is thus generated as

$$E\{\mathbf{y}^{ab}\} = \mathbf{P}^{ab}\lambda = [p_b \mathbf{T}_{pr} \mathbf{W}_m^{-1}\mathbf{R}^{ab}\mathbf{W}_n \mathbf{Q}^a \mathbf{D}^a]\lambda \qquad (7.8)$$

where $\mathbf{D}^a$ is a diagonal matrix containing the probabilities that an emission from a source pixel will strike the a^{th} Ge detector, $\mathbf{Q}^a$ effects a cone-beam mapping of the 3-D source space with the a^{th} Ge detector as the focal point, $\mathbf{H}^{ab} = \mathbf{W}_d^{-1}\mathbf{R}^{ab}\mathbf{W}_s$ contains the probability patterns for the b^{th} Compton scattering angle mapping the cone-beam projections to a polar raster image, $\mathbf{T}_{pr}$ effects a polar -to- rectangular transformation of the polar image, and p_b weights that image with the probability of the b^{th} Compton scattering angle occurring.

A step-by-step diagram of this approach for the forward projection of a 3-D source is shown in figure 7.14. The steps are followed in reverse order to compute the back projections. The result is a dramatic savings in storage requirements and computational demands. Having specified the implementation of forward/back projections for a single detector, the forward/back projections for the other detectors is handled in identical fashion. For each Ge detector, the key is to use the data in the

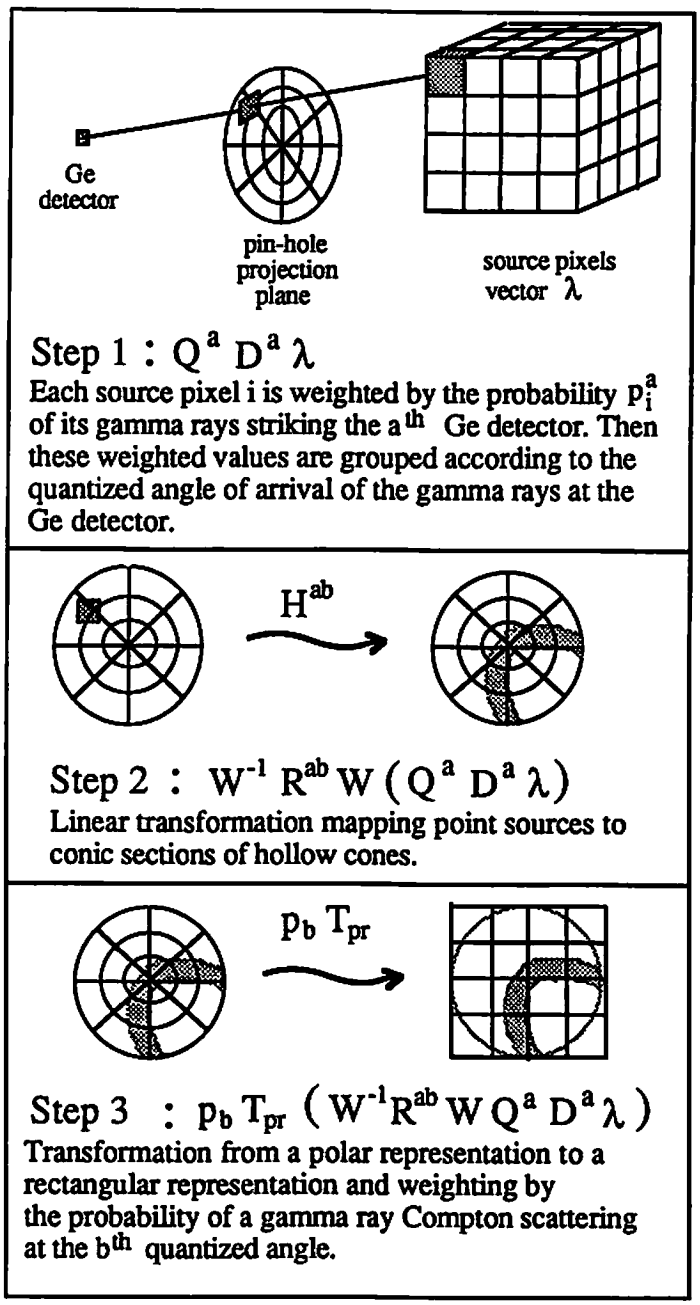

Figure 7.14
A 3-step implementation of a forward projection of the source via the matrix factorization.

largest possible circular region centered at the camera point directly fronted by the Ge detector. The data from each Ge detector and each Compton scattering angle can be processed independently of the other data until the back-projections are finally combined to update the source pixel estimates. This has implications for parallel processor implementation in that all operations except the final updating of the estimate can be performed in parallel with the number of processors equal to the number of detectors times the number of Compton scattering angles or any fraction thereof.

7.4.3 Reconstruction Approaches

In the following we show how the factorization described above may be exploited to develop computationally efficient reconstruction algorithms. Firstly we show how the two stage approaches described in [8],[9],[69] result directly from the two stage factorization in section 7.4.1; secondly we develop a single stage maximum likelihood reconstruction based on the system matrix factorization in section 7.4.2.

7.4.3.1 Two Stage Estimation Approach

The two stage reconstruction approach is based on the matrix factorization of section 7.4.1. A two-stage linear reconstruction approach first estimates the set of 2-D inverted cone beam projections $\hat{\mathbf{x}}$ as an approximate solution to the linear system $\mathbf{y} = \mathbf{A}\mathbf{x}$. Then the source estimate $\boldsymbol{\lambda}$ is formed as a solution of the linear system $\hat{\mathbf{x}} = \mathbf{B}\boldsymbol{\lambda}$. Methods which have been investigated for the first stage of this approach include linear estimators using the ART algorithm [11] and SVD based reconstruction and a maximum likelihood estimator [35]. The second stage requires

reconstruction from a set of cone beam projections and may be performed using filtered backprojection [20] or a second ART procedure [8],[69].

The properties of the two stage estimators are determined by the noise in the data, the rank of the two parts of the factored matrix and the specific algorithm used. Due to the Poisson nature of the data, the linear system $\mathbf{y} = \mathbf{AB}\lambda$ is in general inconsistent, a least squares solution which minimizes $||\mathbf{y} - \mathbf{AB}\lambda||$ is an appropriate linear estimator. Consider the performance of a two stage linear reconstruction approach using a linear least squares estimator at each stage. This method achieves an overall least squares solution under the following condition. For **A**:(m)x(r) and **B**: (r)x(n), if (m) > (r) ≥ (n) and **B** and **A** are of rank of at least (n) then the product of the least squares pseudo-inverses $\mathbf{B}^I$ and $\mathbf{A}^I$, of **B** and **A** respectively, is equal to the pseudo-inverse of **P**, i.e.

$$\mathbf{P}^I = (\mathbf{AB})^I = \mathbf{B}^I\mathbf{A}^I. \tag{7.9}$$

This is easily shown using the singular value decompositions for **A** and **B**.

In practice it is difficult to form a true least squares estimate due to the large dimensions of the matrices **A** and **B**. Instead, a positively constrained version of ART has been used [11]. This method differs in two ways from the least squares approach. Firstly, it is well known that ART converges only for consistent sets of equations. Since for our system the equations are inconsistent, the algorithm will not converge; however, this problem may be overcome by using a stopping criterion to terminate the iterations based on a comparison of the statistics of the noise in the data and those of the residual error [35]. Secondly, the use of the positivity constraint makes the estimator nonlinear, and hence theoretical analysis of the estimator properties becomes very difficult.

The advantage of a two stage approach is that the reconstruction of the cone beam projections is performed separately for each Ge detector. This problem is ideally suited to implementation on a set of parallel processors and hence, since the dominant computational cost is this first stage, a two stage approach has obvious computational advantages. In the following section we exploit the matrix factorization, described above, to achieve similar computational advantages in a single stage 3D ML reconstruction algorithm.

7.4.3.2 The Maximum Likelihood Reconstruction Method

In contrast to the linear reconstruction approach discussed in the previous section, the maximum likelihood (ML) approach uses a probabilistic description of the data generating process. In the sections which follow, we describe a computationally frugal implementation of a one-stage 3-D ML reconstruction approach. The implementation is based upon the probabilistic factorization of the system matrix as described in section 7.4.2. The formulation is based on the Poisson model for emissions from a radioactive source as discussed in section 2.3.1.3. Let the vector containing the unknown means of pixels in the 3-D region of interest be denoted by $\boldsymbol{\lambda}$, the data vector by $\mathbf{y}$, and the probability that an emission from source pixel j would generate a count at data element i by $\mathbf{P}_{ij}$. The probability distribution of the data conditioned upon the unknown image $\boldsymbol{\lambda}$ thus has the form

$$f(\mathbf{y}\,|\,\boldsymbol{\lambda}) = \prod_i \frac{(\sum_j \mathbf{P}_{ij}\lambda_j)^{y_i}}{(y_i)!}\, e^{-\sum_j \mathbf{P}_{ij}\lambda_j} \tag{7.10}$$

The maximum likelihood approach determines the vector $\boldsymbol{\lambda}$ which maximizes $f(\mathbf{y}\,|\,\boldsymbol{\lambda})$ or equivalently maximizes any monotonic function of $f(\mathbf{y}\,|\,\boldsymbol{\lambda})$ such as $\ln f(\mathbf{y}\,|\,\boldsymbol{\lambda})$.

Dempster, Laird and Rubin's expectation-maximization (EM) algorithm [18] applied to PET reconstruction by Shepp and Vardi [65]. has the same form for all emission imaging systems. At iteration k the estimate of the unknown image vector λ^k is updated according to the formula

$$\lambda^{k+1} = \frac{\lambda^k}{\sum_q \mathbf{P}_{qk}} \sum_i \frac{\mathbf{P}_{ij} \mathbf{y}_i}{\sum_k \mathbf{P}_{ik} \lambda_k^k} \tag{7.11}$$

Implementation begins with an initial non-zero estimate for λ, usually a constant image, and requires four steps:

(1) forward projection of the present source estimate $\bar{\mathbf{y}} = \mathbf{P}\lambda$

(2) normalization of the data by the forward projection $\hat{\mathbf{y}}_i = \mathbf{y}_i / \bar{\mathbf{y}}_i \qquad i+1,..m$

(3) backprojection of the normalized data $\mathbf{t}^k = \mathbf{P}^t \hat{\mathbf{y}}$

(4) updating of the image estimate $\lambda_j^{k+1} = (\lambda_j^k \; \mathbf{t}_j^k)/(\sum_q \mathbf{P}_{qj}) \qquad j=1,..n$

It can easily be shown that the likelihood function is concave over $\Omega=\{\lambda \in R^n \; , \lambda_i \geq 0 \; , \; i=1,2,..n \}$ and consequently any local optimum is a global optimum. Furthermore, a necessary and sufficient condition for the uniqueness of the maximum likelihood estimate is that the Hessian of the negative likelihood function (i.e. the matrix of partial derivatives of $-\ln f(\mathbf{y}|\lambda)$, is strictly positive definite for all $\lambda \in \Omega$. Taking partial derivatives of the log of equation (7.10), the Hessian can be shown to equal

$$\mathbf{H(k,l)} = \sum_{i=1}^{m} \frac{y_i \mathbf{P}_{ik} \mathbf{P}_{il}}{(\sum_{j=1}^{n} \mathbf{P}_{ij} \lambda_j)^2} \; . \tag{7.12}$$

From the properties of positive definite matrices it can further be shown that a necessary condition for positive definiteness is that $\mathbf{P}$ be of rank at least n; a

sufficient condition is that **P** is of rank at least n and the data vector **y** is strictly positive. A single necessary and sufficient is condition is that the $n \times m$ matrix

$$\hat{\mathbf{P}}(i,l)=\mathbf{P}_{il}\left[y_i^{1/2}/(\sum_{j=1}^{n}\mathbf{P}_{ij}\lambda_j)\right] \tag{7.13}$$

be of rank at least n for all $\lambda \in \Omega$. All of these conditions are clearly difficult to test in practice. However, since $m >> n$ and no two rows of **P** are identical it is very likely that **P** is of rank at least n. In addition, an examination of the necessary and sufficient condition indicates that provided a significant number of the data values are non zero, it is very likely that a unique ML estimate exists. This is an important property, since we can predict that the ML method applied to the electronically collimated system will not suffer from the instability which is commonly encountered in conventional SPECT [34] caused by the ill conditioned nature of non-unique solutions to high dimensional reconstruction problems. This prediction is verified in our simulations in which we have seen no evidence of ill conditioning manifested by deterioration of the reconstruction at high iteration numbers.

The two stage reconstruction approach proposed in [35] uses a ML reconstruction of the cone-beam projection images of the Poisson means of the source pixels followed by a least squares reconstruction of the source pixels from the cone-beam projection images. Using the two stage factorization described in section 7.4.1 and the invariance property of the ML estimate, we can derive conditions under which this method results in a true ML solution. The invariance property states that if a vector of parameters $\lambda \in R^n$ can be expressed as a function $\lambda = f(x)$ of some other vector of parameters $x \in R^r$ where $f(\cdot)$ is a unique mapping from R^r to R^n , then $\lambda_{ML} = f(x_{ML})$. Using $E\{y\} = AB\lambda$ we may find a ML solution x_{ML} to the subsystem $E\{y\} = Ax_{ML}$. The ML estimate of λ may then be found from x_{ML} if and

only if we can express the relationship between λ and $\mathbf{x}$ as some function $\lambda = f(x)$. For this imaging system, the relationship between $\mathbf{x}$ and λ is given by the linear transformation $\mathbf{x} = \mathbf{B}\lambda$. Therefore, a necessary and sufficient condition under which λ_{ML} may be found from $\mathbf{x}_{ML}$, is that the inverse of $\mathbf{B}$ exists, i.e. $\mathbf{B}$ is square and of full rank. In this case $\lambda_{ML} = \mathbf{f}(\mathbf{x}_{ML}) = \mathbf{B}^{-1}\mathbf{x}_{ML}$. This condition is not satisfied for the system described above and therefore we can conclude that the two stage approach employing maximum likelihood estimation in the first stage does not yield a true maximum likelihood estimate.

For the electronically collimated system we have implemented the single stage ML algorithm for the configuration, described in the previous section, based on 320 germanium detectors. Each Ge detector generates eight 90x90 2-D photon count images, one for each quantized Compton scattering angle level. The data vector $\mathbf{y}$ is formed by concatenating the images for each detector and each scatter angle (figure 7.11) into a single vector of 20,736,000 elements. Since all data pixels have positive integer values of less than 255, this requires one byte per element. The source image is divided into 48^3 pixels which can be concatenated into a vector whose elements take real values and are stored as real*4 variables.

Implementation of the EM algorithm requires both forward and back projection through the matrix $\mathbf{P}$ at each iteration. These forward/back projections were implemented by using the probabilistic factorization of the system matrix as described in section 7.4.2. To demonstrate the feasibility of this approach, we simulated 12.8×10^6 detected photon counts from a 12 cm sphere of activity containing a 'hot' region and a 'cold' region, of relative intensities 1, 2, and 0 respectively. The Ge detectors were placed a distance of 11 cm from the center of the sphere and the gamma camera 5 cm behind the Ge detectors. The image was reconstructed on a

48^3 pixel cube with cubic pixels 0.33 cm on each side. A Fortran coding of this approach required on the order of 9 hours of cpu time per iteration on a Sun 3/110 workstation with a floating point accelerator.

7.4.4 Speeding the Convergence by Iterating Upon Subsets of the Data

The back projection of so many normalized data values results in a very smooth multiplier image for the updating stage of the algorithm (step 4 above). Convergence can thus be very slow. This led us to examine a unique modification to the EM algorithm. To increase the speed the convergence of the algorithm, the data is broken into 16 subsets, one subset for each Ge detector. Each subset contains the data associated with a single detector at all 20 viewing angles. Each subset is treated as the complete data set for one iteration of the EM algorithm. These 16 data subsets are cycled through sequentially until the likelihood, based upon all the data, no longer increases. The 16 data subsets are then paired and the EM algorithm continued in similar fashion cycling through the 8 data subsets. When the likelihood based on all the data is no longer increased, iterations based upon all the data are performed until sufficient convergence has been achieved.

To evaluate the approaches presented in this section, we simulated the data generated by an electronically collimated imaging system by forward projecting a spherical computer generated source through the linear transformation representing the system and generated independent Poisson random variables with mean equal to the forward projection. Figure 7.15 shows four data images collected from a single Ge detector at four different Compton scattering angles. The 3-D source image is a single point containing a concentration of radiolabeled material. The upper left image contains detected counts which Compton scattered within the lowest

recorded angular interval. The subsequent images are for increasingly larger recorded Compton scattering angles. The reconstructions of the source after 6 and 30 iterations using the complete data at each iteration are shown in figures 7.16 and 7.17 respectively. Further iterations past 30 resulted in very little change in the likelihood function and no noticeable image improvement. The increase in the likelihood from iteration 6 of the EM ML algorithm through iteration 30 was significant and resulted in a marked improvement in the reconstruction of the spherical 3-D source. For comparison, in figure 7.18 we show the result of the EM ML algorithm iterations using subsets of the data after 6 iterations through all data subsets. The reconstructed images in figures 7.17 and 7.18 are of comparable quality. In fact, the likelihood, based upon of the data, of figure 7.18 is higher than that of figure 7.17. Figure 7.16, which is the result of 6 EM ML algorithm iteration based on all the data lacks detail, particularly at the edges. This is an indication that convergence of the likelihood was not yet achieved.

The convergence rates for the EM ML algorithm based on all the data and the EM ML algorithm based on the particular subsets of the data suggested above are shown in figure 7.19. Plotted is the likelihood, based on all the data, at each of 10 iterations of the respective algorithms. Following the argument above, a unique global maximum should exist. Since intial ML iterations based upon data subsets are to followed by ML iterations based on all the data, both schemes will converge to the same solution, differing only in the path by which that solution is reached. Since the rate of convergence becomes increasingly slow as the optimal solution is approached, in practice the iterations must be terminated prior to convergence. It is advantageous to employ a scheme in which initial convergence is fast such as that described above. Since the computation cost for one complete iteration using all the data is approximately equal to the cost of performing one complete iteration

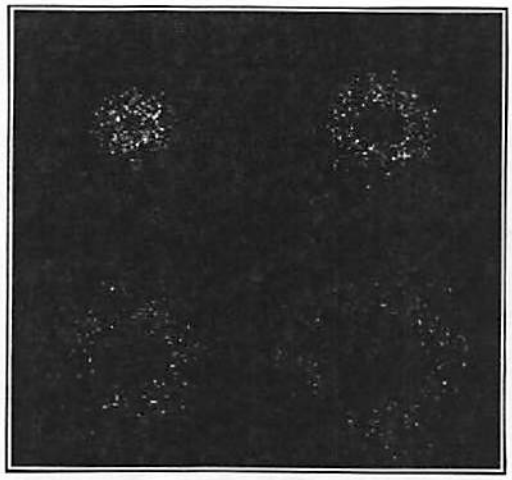

Figure 7.15
Four data images collected from a single Ge detector at four different Compton scattering angles.

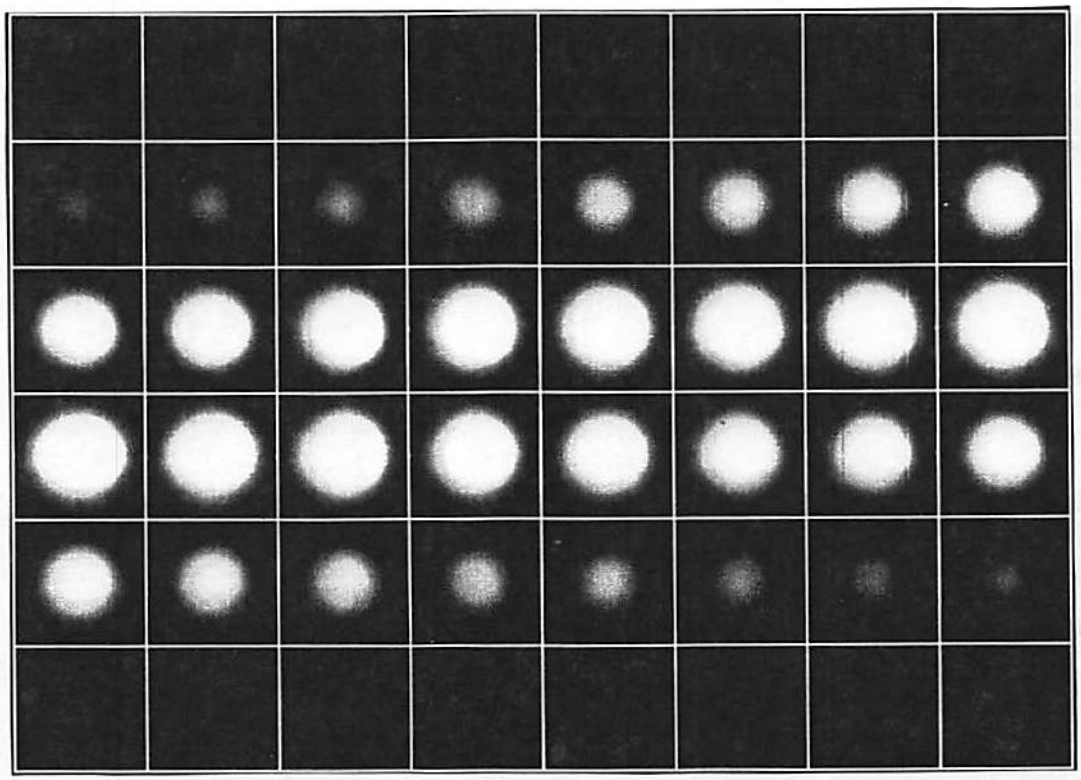

Figure 7.16
Reconstruction of sphere source from 12.8 million counts spread over 20.7 million data values using 6 iterations of the EM algorithm.

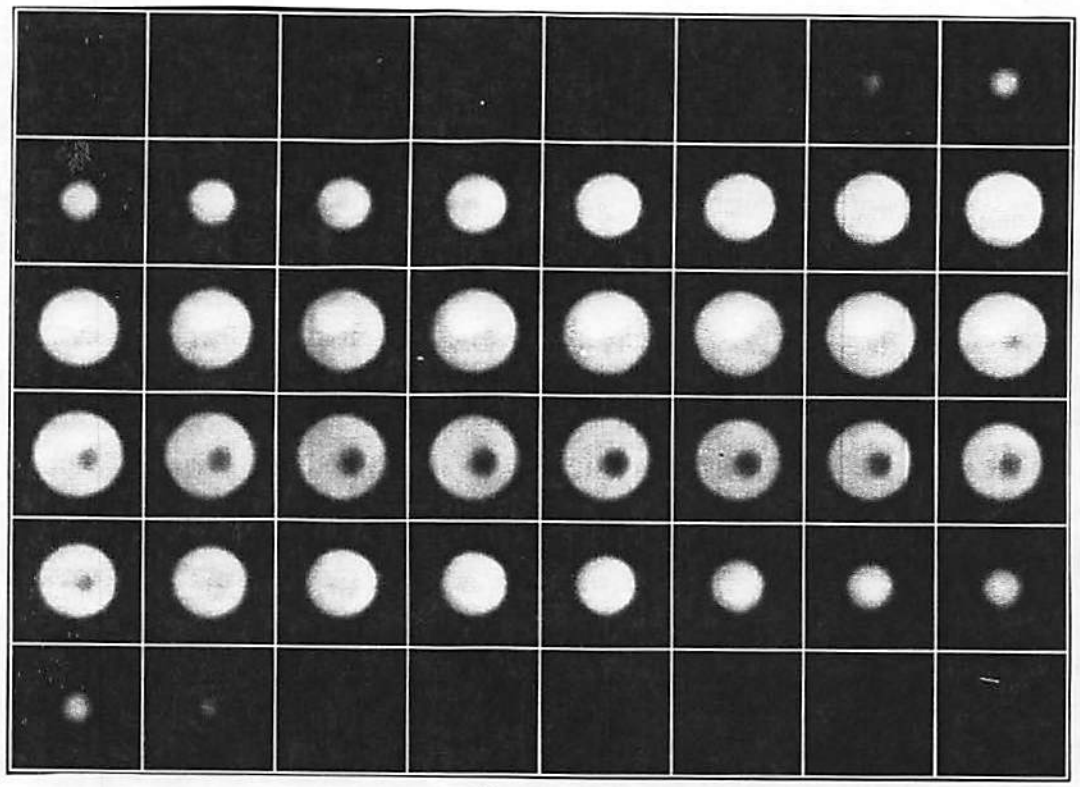

Figure 7.17
ML reconstruction from 12.8 million counts over 20.7 million data values using 30 EM iterations.

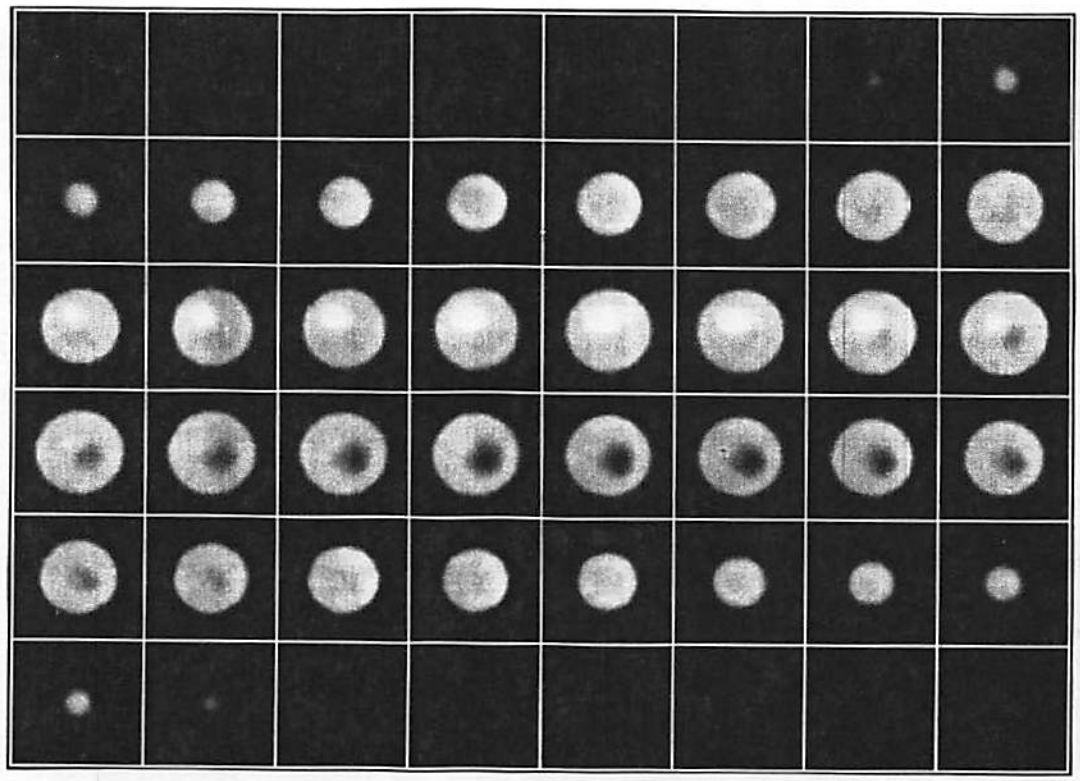

Figure 7.18
Reconstruction of sphere using 4 EM iterations based on data subsets and 2 iterations based on all data.

based on all the subsets of data, this scheme offers a potential improvement in the convergence rate of approximately a factor of five.

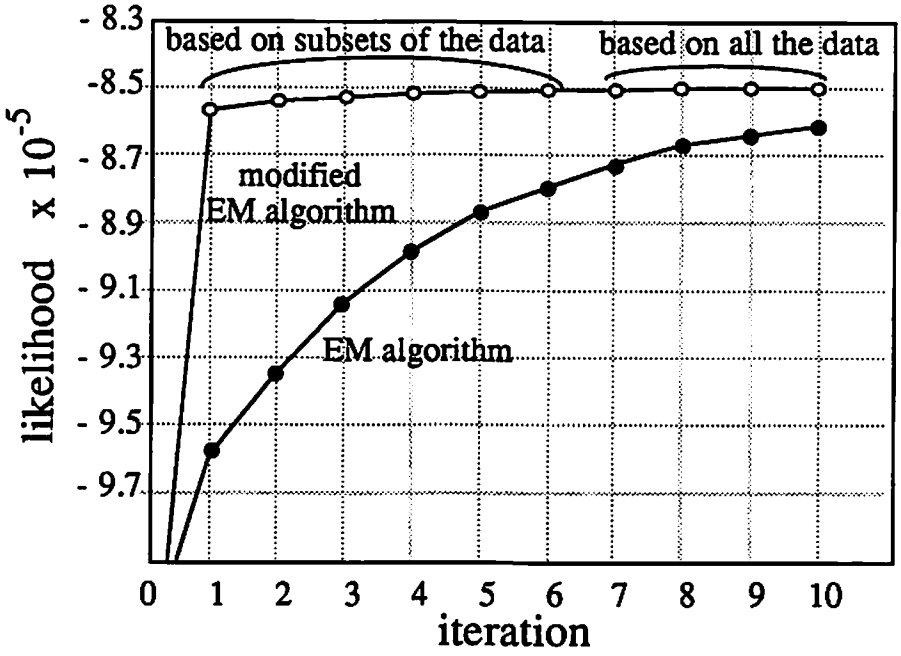

Figure 7.19
Likelihood based on all the data vs. iteration of the ML reconstruction algorithm for the EM algorithm and the modified EM algorithm based on subsets of the data.

What has been achieved is a dramatic increase in the speed of convergence towards the ML solution in the initial stages of the algorithm. This cannot be maintained if iterations based upon subsets of the data are continued. Convergence to the ML solution cannot be achieved solely by iterations based upon subsets of the data. This is why a few iterations based upon all of the data are required to achieve acceptable convergence of the algorithm. This approach did result in a substantial reduction in the overall computation time. Where 30 iterations of the EM algorithm based upon all of the data required 14 days of CPU time, the approach shown here

required but 48 hours of CPU time and resulted in greater convergence of the likelihood. This is still a substantial computational investment. However, without the probabilistic modeling based on matrix factorizations and without the ART type modification to the EM ML algorithm, reconstruction of radiolabeled sources from data obtained by this system would require either prohibitively more computation time or simplifications and approximations which are not born out by the data generating mechanism of the emission imaging system. What has been achieved is a direct, iterative, 3-D maximum likelihood reconstruction for a system with a spatially variant point source response.

Chapter 8

Conclusions

In this dissertation we have addressed the topic of Bayesian estimation for image restoration and reconstruction as a probabilistic method of incorporating qualitative information regarding the smoothness of images. We began by examining maximum likelihood estimation, Bayesian estimation, and regularization procedures. Then we embraced the concept of a Markov random field as a natural extension of a Markov process to multi-dimensions. The Hammersly-Clifford theorem [6] revealed the equivalence between Markov random fields and Gibbs distribution. We discussed algorithms which have been applied for Bayesian estimation. In [22], Geman and Geman applied the simulated annealing algorithm introduced by Kirkpatrick [43] to seek globally optimal MAP estimates based on Gibbs priors. They pointed out the limitations of this approach for applications, such as emission tomography, in which the blurring function is not resticted to local interactions. This is where we began.

First, we confirmed the need to incorporate qualitative information regarding the local smoothness of images. As a computationally simple approach to this problem we suggested a statistic based stopping criterion for iterative reconstruction algorithms such as iterative least squares and maximum likelihood. The motivating factor in this approach was to regularize the solution by effecting a trade-off between the maximum likelihood solution and a maximally smooth uniform image. This approach was shown to outperform maximum likelihood image estimation while decreasing the required amount of computation.

Next we formulated the image restoration and reconstruction problems in the complete/incomplete data framework of Dempster, Laird, and Rubin and derived an

expectation-maximization algorithm for MAP estimation based on a particular Gibbs prior. Using a similar optimization approach, we derived an algorithm for MAP estimation of images based on any Gibbs prior. This algorithm combined the complete/incomplete data formulation of the EM approach, section 3.4.1.2, coordinate ascent [52], the optimal step-size of a maximum likelihood algorithm derived along parallel lines, and a method of decreasing the step size if necessary. Each iteration of this algorithm begins with some computations performed over the whole image. This is followed by sequential updating of the image pixels based upon local calculations. This two step algorithm requires less computation than the ICM algorithm because after the first step, the interaction between the data and image pixels which are not neighbors of the image pixel being updated is decoupled. In addition, the computational burden of performing a line search is omitted. These two GEM MAP algorithms were presented as deterministic alternatives to simulated annealing. They require far less computation but, unlike simulated annealing, can only be shown to attain locally optimal solutions. We used these algorithms to compare three different Gibbs priors. The third prior was formulated to overcome an undesireable property of the prior suggested by Geman and McClure in [23]. The available flexibility in defining the potential functions of a Gibbs prior is an area which has yet to be completely explored.

The Gibbs functions which we use as prior distributions for images have a very important parameter which specifies the degree to which the modes of the distribution are accentuated. In practice, this parameter is generally selected on the basis of trial and error. Since this parameter is synonymous with the inverse of a regularization parameter we examined accepted methods of selecting a regularization parameter. Furthur, we examined the single published method of selecting a Gibbs prior parameter introduced by Geman and McClure in [23]. Then we

formulated a statistic based approach to selection of the Gibbs prior parameter and we evaluated its potential for selecting the optimal parameter with respect to L_1 and L_2 restoration error. The resulting solution was set in a control systems framework and an algorithm for use in conjunction with any deterministic MAP algorithm for simultaneous parameter selection and image restoration or reconstruction was presented. This method requires updating of the prior parameter as the iterative restoration or reconstruction proceeds. It produced very good results and the complete restoration/parameter selection algorithm converged quite quickly.

We addressed some practical problems in emission tomography: reducing computation requirements and attenuation compensation. Emission tomography presents a 3-D reconstruction problem, the solution of which suffers from being extremely computationally intensive. By using an intermediate polar pixel representation of the unknown 3-D distribution of radioisotope we were able to decrease the computational demand of implementing iterative reconstruction algorithms. The computational savings was shown to be proportionate to the non-localness of the system matrix: the more spread out the point source response - the greater the computational advantage of this approach. In order to achieve the greatest possible resolution from an emission imaging modality, it becomes necessary to model the point source response of a system as accurately as possible. As the trend in reconstruction swings in this direction, the computation required to implement popular reconstruction algorithms increases. Methods of reducing those computations while retaining the benefit of highly accurate modeling are therefore becoming more and more important.

Finally, there are many new emission imaging systems being proposed. One system under development at this university is based upon electronic collimation.

This system incorporates the measurement of the Compton scattering angle of detected photons. As such, the point response function is spatially variant and the dimension of the reconstruction is unusually large, 48^3 pixels to be estimated from 20.7×10^6 data values. We performed a probabilistic modeling of the data generating mechanism of the system and effect an iterative maximum likelihood reconstruction. A probabilistically based factorization of the system matrix which was introduced allowed reduction of the computational demands of a class of iterative algorithms to be reduced by several orders of magnitude. An 'algebraic reconstruction technique' type modification to a well known ML algorithm was introduced producing an increase in the likelihood convergence speed by a factor greater than five.

References

[1] K.J. Astrom, U. Borrison, L.Ljung, B. Wittenmark; "Theory and Applications of Self-Tuning Regulators;" *Automatica*, Vol. 13, pp 457-476, 1977.

[2] H. Barrett, W. Swindell; *Radiological Imaging, Volume 1* ; Academic Press, New York 1981

[3] P. Bell, W. McClain, D. Ross; "A Computer-Based Scanning System for Nuclear Medicine;" *IEEE Transactions on Nuclear Science*, Vol. 20, Feb. 1972.

[4] J. Berger; *Statistical Decision Theory and Bayesian Analysis*; Springer-Verlag, New York, 1985.

[5] J. Besag; "On the Statistical Analysis of Dirty Pictures;" *Journal of the Royal Statistical Society*, No. 3, 1986.

[6] J. Besag; "Spatial Interaction and the Statistical Analysis of Lattice Systems;" *Journal of the Royal Statistical Society*, series B, Vol. 34, 1972.

[7] A. Blake, A. Zisserman; *Visual Reconstruction*; MIT Press, Cambridge, Massachusetts, 1987.

[8] R. Brechner, M. Singh, R. Leahy; "Computer Simulated Studies of Reconstruction with an Electronically Collimated Camera for SPECT;" *IEEE Transactions on Nuclear Science,* NS-34, pp 369-373, 1987.

[9] R. Brechner, M. Singh; "Reconstruction of Electronically Collimated Images Obtained from Single Photon Emitters using a Spherical System of Coordinates;" *IEEE Transactions on Nuclear Science,* NS-33, pp 583-586, 1986.

[10] T. Budinger, G. Gullberg; "3-D Reconstruction in Nuclear Medicine Emission Imaging;" *IEEE Transactions on Nuclear Science*, Vol. 24, June 1974.

[11] Y. Censor; "Finite Series Expansion Reconstruction Methods;" *Proceedings of the IEEE* , Vol 71, No. 3, pp 409-419, March 1983.

[12] L. Chang; "Attenuation Correction and Incomplete Projection in Single Photon Emission Computed Tomography;" *IEEE Transactions on Nuclear Science,* Vol. 26, April 1979.

[13] R. Chellappa, R. Kashyap; "Digital Image Restoration using Spatial Interaction Models;" *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. ASSP-30, No. 3, pp 461-472, June 1982.

[14] H. Cramer; *Mathematical Methods of Statistics*; Princeton University Press, 1946.

[15] P. Craven, G. Wahba; "Smoothing Noisy Data with Spline Functions;" *Numerische Mathematik*, Vol. 31, pp 377-403, 1979.

[16] G.R. Cross, A.K. Jain; "Markov Random Field Texture Models;" *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-5, No. 1, pp 25-39, January 1983.

[17] A.R. Davies; " On the Maximum Likelihood Regularization of Fredholm Equations of the First Kind;" in *Treatment of Integral Equations by Numerical Methods*, edited by C. Baker and G.F. Miller, pp 95-105, Academic Press, 1982.

[18] A.P. Dempster, N.M. Laird, D.B. Rubin; "Maximum Likelihood from Incomplete Data via the EM Algorithm;" *Journal of the Royal Statistical Society*, series B, Vol. 39, 1977.

[19] M. Ekstrom, O. Rater; "Computer Processing of Radigraphic Images with the FFT;" *IEEE Transactions on Nuclear Science*, Vol. 19, February 1972.

[20] L. Feldkamp, L. Davis, J. Kress; "Practical Cone-beam Algorithm;" *Journal of the Optical Society of America A*, Vol-1 No. 6, pp 612-619, 1984.

[21] E. Gamble, T. Poggio; "Visual Integration and Detection of Discontinuities: the Key Role of Intensity Edges;" A.I. Labratory Memo No. 970, MIT A.I. Lab, October 1987.

[22] S. Geman, D. Geman; "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images;" *IEEE Transactions on Pattern Analysis and Machine Intelligence*, No. 6, November 1984.

[23] S. Geman, D. McClure; "Bayesian Image Analysis: an Application to Single Photon Emission Tomography;" *Proceedings Statist. Comput. Sect.*, American Statistical Association, 1985

[24] S. Geman, D. Geman; "Statistical Methods for Tomographic Image Reconstruction;" *Proceedings of the 46th Session of the ISI, Bulletin of the ISI*, Vol. 52, 1987.

[25] S. Goldman, L. Albrecht, B. Wheaton; "An Analysis of Digitally Processed Readouts from Scintillation Cameras and Scanners;" *IEEE Transactions on Nuclear Science* Vol. 17, February 1970.

[26] G. Golub, M. Heath, G. Wahba; "Generalized Cross-Validation as a Method for Choosing a Good Ridge Parameter;" *Technometrics*, Vol. 21, No. 2, pp 215-223, May 1979.

[27] R. Gonzalez, P. Wintz; *Digital Image Processing*; Addison-Wesley Publishing Co.; Reading, Massachusetts; 1977.

[28] G.C. Goodwin, P.J. Ramadge, P.E. Caines; "Discrete-Time Multivariable Adaptive Control;" *IEEE Transactions on Automatic Control*, Vol. AC-25, No. 3, pp 449-456, June 1980.

[29] R. Gordon; " A Tutorial on ART (Algebraic Reconstruction Technique);" *IEEE Transactions on Nuclear Science*, Vol. 24, June 1974.

[30] U. Grenander; *Abstract Inference*; John Wiley & Sons, 1981.

[31] C.J. Harris, S.A. Billings; *Self-Tuning and Adaptive Control: Theory and Applications*; Peter Peregrinus LTD on behalf of the Institute of Electrical Engineers, London and New York, 1981.

[32] H. Hart, Z. Liang; "Bayesian Processing in Two Dimensions;" *IEEE Transactions on Medical Imaging*, Vol. MI-6, No. 3, pp 201-208, September 1987.

[33] T. Hebert, R. Leahy; "A Generalized EM Algorithm for 3-D Bayesian Reconstruction from Poisson Data using Gibbs Priors;" to appear in *IEEE Transactions on Medical Imaging*, June 1989.

[34] T. Hebert, R. Leahy, M. Singh; "Fast MLE for SPECT using an Intermediate Polar Representation and a Stopping Criterion;" *IEEE Transactions on Nuclear Science, Vol. NS-34, pp 615-619*, February 1988.

[35] T. Hebert, R. Leahy, M. Singh; "ML Reconstruction for a Prototype Electronically Collimated Single Photon Emission System;" *SPIE Medical Imaging Conference Proceedings*, pp 77-83, February 1987.

[36] G.T. Herman; *Image Reconstruction from Projections*; Academic Press, 1980.

[37] A. Hoerle, R. Kennard; "Ridge Regression: Biased Estimation for Non-Orthogonal Problems;" *Technometrics*, 12, pp 55-67, 1970.

[38] J. Jones; "Current Problems in Ultrasound Impediography;" *Proceedings of Seminar on Ultrasound Tissue Characterization, National Bureau of Standards*; Gaithersburg, MD, pp 253-257, May 28-30 1975.

[39] A. Kak, F. Fry; "Acoustic Impedance Profiling: an Analytical and Physical Model Study;" *Proceedings of Seminar on Ultrasound Tissue Characterization, National Bureau of Standards*, Gaithersburg, MD; pp 231-251, May 28-30 1975.

[40] R. Kashyap, R. Chellappa; "Estimation and Choice of Neighbors in Spatial Interaction Models of Images;" *IEEE Transactions on Information Theory*, Vol. IT-29, No. 1, pp 60-72, January 1983.

[41] L. Kaufman; "Implementing and Accelerating the EM Algorithm for Positron Emission Tomography;" *IEEE Transactions on Medical Imaging* , No. 1, March 1987.

[42] K.J. Kearfott; "Comment: Practical Considerations;" *Journal of the American Statistical Association*, March 1985, pp. 26-28.

[43] S. Kirkpatrick, C. Gelatt, M. Vecchi; "Optimization by Simulated Annealing;" *Science*, Vol. 22, pp 671-680, 1983.

[44] J. Kormylo, J. Mendel; "Maximum-Likelihood Seismic Deconvolution;" *IEEE Transactions on Geoscience & Remote Sensing*, Vol. GE-21, No. 1, pp 72-82, January 1983.

[45] D.T. Kuan, A.A. Sawchuk, T. Strand, P. Chavel; "Adaptive Noise Smoothing Filter for Images with Signal Dependent Noise;" *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 7, No. 2, pp 165-177, March 1985.

[46] M. Kung, B. Womack; "Discrete Time Adaptive Control of Linear Dynamic Systems with a two-Segment Piecewise-Linear Asymmetric Nonlinearity;" *IEEE Transactions on Automatic Control*, Vol. AC-29, No. 2, pp 170-172, February 1982.

[47] P. van Laarhoven, E. Aarts; *Simulated Annealing: Theory and Applications*; D. Reidel Publishing Company, 1986.

[48] E. Levitan, G.T. Herman; "A Maximum A Posteriori Probability Expectation Maximization Algorithm for Image Reconstruction in Emission Tomography;" *IEEE Transactions on Medical Imaging*, Vol. MI-6, No. 3, pp 185-192, September 1987.

[49] Z. Liang, H. Hart; "Bayesian Image Processing of Data from Constrained Source Distributions-Non-Valued, Uncorrelated and Correlated Constraints;" *Bulletin of Mathematical Biology*, Vol. 49, pp 51-74, 1987.

[50] B. Lindgren; *Statistical Theory*; Macmillan Publishing Co., New York, Third Edition, 1976.

[51] J. Llacer, E. Veklerov, E. Hoffman; "On the Convergence of the MLE Method of Tomographic Image Reconstruction;" *SPIE Medical Imaging Conference Proceedings*, February 1987.

[52] D. Luenberger; *Linear and Non-Linear Programming*; Addison-Wesley Publishing Co., Readings Mass., Second Edition, 1984.

[53] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller; "Equation of State Calculations by fast Computing Machines;" *Journal of Chemical Physics*; Vol. 21, No. 6, pp 1087-1092; June 1953.

[54] M. Miller, D. Snyder, S. Moore; "An Evaluation of the Use of Sieves for Producing Estimates of Radioactivity Distributions with the EM Algorithm for PET;" *IEEE Transactions on Nuclear Science*; Vol. NS-33, 1986.

[55] D. Mitra, F. Romeo, A. Sangiovanni-Vincentelli; "Convergence and Finite Time Behavior of Simulated Annealing;" *Adv. Appl. Prob.*, Vol. 18, pp 747-771, 1986.

[56] G.E. Noether; *Elements of Nonparametric Statistics*; John Wiley & Sons, Inc., New York, 1967

[57] R.L. Obeinchain; "Good and Optimal Ridge Estimators;" *Annals Statistics*, Vol. 6, pp 1111-1121, 1978.

[58] C.R. Rao; *Advanced Statistical Methods in Biometric Research;* Hafner Press, N.Y., Second Reprint, 1974.

[59] E. Roberts, T. Stapinski, A. Rodgers; "Photon Counting Detectors for the Fuse/Lyman Satellite Telescope;" *Journal of the Optical Society of America A*, Vol. 3, No. 12, pp 2146-2150, December 1986.

[60] A. Rockmore, A. Macovski; "A Maximum Likelihood Approach to Emission Image Reconstruction from Projections;" *IEEE Transactions on Nuclear Science,* Vol. 23, 1976.

[61] A. Rosenfeld, A. Kak; *Digital Picture Processing I & II*; Academic Press, 1982.

[62] R. Serfling; *Approximation Theorems of Mathematical Statistics*; John Wiley & Sons, 1980.

[63] L. Shepp, Y. Vardi, et al; "Maximum Likelihood PET with Real Data;" *IEEE Transactions on Nuclear Science*, Vol. NS-31 No. 2, pp 910-913, April 1984.

[64] L. Shepp, Y. Vardi; "Maximum Likelihood Reconstruction in Positron Emission Tomography;" *IEEE Transactions on Medical Imaging*, Vol. MI-1, pp 113-122, 1982.

[65] L. Shepp, Y. Vardi, L. Kaufman; "A Statistical Model for Positron Emission Tomography;" *Journal of the American Statistical Association*, Vol. 80, No. 389, pp 8-20, March 1985.

[66] T. Simchony, R. Chellappa, Z. Lichtenstein; "Graduated Non-Convexity Algorithm for Image Estimation using Compound Gauss Markov Field Models;" Signal and Image Processing Institute Report No. 128, 1988, University of Southern California, Los Angeles, California.

[67] M. Singh; "An Electronically Collimated Gamma Camera for Single Photon Emission Computed Tomography. Parts 1 and 2;" *Medical Physics* , Vol 10, No. 4, pp 421-427, Jul/Aug 1983

[68] M. Singh, D. Doria; "Germanium-scintillation Coincidence Detection Studies for Imaging Single Photon Emitters;" *IEEE Transactions on Nuclear Science*, Vol NS-31, pp 594-598, 1984.

[69] M. Singh, D. Doria; "Single Photon Imaging with Electronic Collimation;" *IEEE Transactions on Nuclear Science*, Vol NS-32, pp 843-848, 1985.

[70] M. Singh, R. Brechner, R. Leahy, T. Hebert; "Noise Propagation in Electronically Collimated Single Photon Imaging;" *IEEE Transactions on Nuclear Science*, Vol NS-35, pp 772-777, 1988.

[71] D. Snyder, M. Miller; "The Use of Sieves to Stabilize Images Produced with the EM Algorithm for Emission Tomography;" *IEEE Transactions on Nuclear Science*, Vol. NS-23, No. 5, pp 3864-3870, October 1985.

[72] H. Sorenson; *Parameter Estimation: Pronciples and Problems*; Marcel Dekker, New York; 1980.

[73] O.N. Strand; "Theory and Methods Related to the Singular-Function Expansion and Landwebers Iteration for Integral Equations of the First Kind;" *SIAM Journal of Numerical Analysis*, Vol. 12, No. 4, pp 798-825, September 1974.

[74] C.M. Theobald; "Gneralizations of Mean Square Error Applied to Ridge Regression;" *Journal of the Royal Statistical Society*, Series B, Vol. 36, pp 103-106, 1974.

[75] A. Tikhonov, V. Arsenin; *Solutions of Ill-Posed Problems*; Winston and Sons, Washington, D.C., 1977.

[76] D.M. Titterington; "Common Structure of Smoothing Techniques in Statistics;" *International Statistics Review*, 53, 2, pp 141-170, 1985.

[77] D.M. Titterington; "General Structure of Regularization Procedures in Image Reconstruction;" *Astronomy and Astrophysics*, 144, pp 381-387, 1985.

[78] D.M. Titterington; "Regularization Procedures in Signal Processing and Statistics;" in *Treatment of Integral Equations by Numerical Methods*, edited by C. Baker and G.F. Miller, pp 214-223, Academic Press, 1982.

[79] H.J. Trussell; "Convergence Criteria for Iterative Restoration Methods;" *IEEE Transactions on Acoustics, Speech and Signal Processing*, Vol. ASSP-31, pp. 129-136, 1983.

[80] E. Veklerov, J. Lacer; "Stopping Rule for the MLE Algorithm Based on Statistical Hypothesis Testing;" *IEEE Transactions on Medical Imaging*, No. 4, December 1987.

[81] G.N. Watson; *Theory of Bessel Functions;* Cambridge, The University Press, 1952.

[82] G. Wahba; "Practical Approximate Solutions to Linear Operator Equations when the Data are Noisy;" *SIAM Journal on Numerical Analysis*, Vol. 14, No. 4, pp 651-667, September 1977.

[83] J. Woods; "Two-Dimensional Discrete Markovian Random Fields;" *IEEE Transactions on Information Theory*, Vol. 18, pp 232-240, March 1972.

[84] C.F.J. Wu; "On the Convergence Properties of the EM Algorithm;" *The Annals of Statistics*, Vol. 11, pp 95-103, 1983.