8-2013

An Investigation of Markov Random Fields for Bayesian Reconstruction of Single Photon Emission Computed Tomography

Lloyd Fredrick Arrowood
arrowood@utk.edu

To the Graduate Council:

I am submitting herewith a dissertation written by Lloyd Fredrick Arrowood entitled "An Investigation of Markov Random Fields for Bayesian Reconstruction of Single Photon Emission Computed Tomography." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Computer Science.

Jens Gregor, Major Professor

We have read this dissertation and recommend its acceptance:

Michael Berry, Charles Collins, Michael Thomason, Jonathan Wall

Accepted for the Council:

Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

An Investigation of Markov Random Fields for Bayesian Reconstruction of Single Photon Emission Computed Tomography

A Dissertation

Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Lloyd Fredrick Arrowood

August 2013

Dedication

This work is dedicated to my wife, Sandra,
my father and the memory of my mother.

Acknowledgments

I would like to express my utmost gratitude to my advisor, Professor Jens Gregor, for his continuous guidance and encouragement, as well as his extraordinary patience, during the course of my research. I would like to thank Professor Michael Berry, Professor Charles Collins, Professor Michael Thomason and Professor Jonathan Wall for serving on my doctoral committee and for their constructive criticism on the dissertation. I would like to thank Dr. Hector Santos-Villalobos and Dr. Philip Bingham for their technical guidance on the calculation and assessment of MTF and CDF curves to determine image quality and the use of their Matlab codes that were extended to perform analyses of 3-D images. Professor Gene Gindi from the State University of New York at Stony Brook and Professor Soo-Jin Lee from Pai Chai University, South Korea provided helpful insight into their research on mechanical priors. Mr. Sanghyeb Lee generated the mouse spleen data sets using the GATE simulation package.

Most importantly, I would like to thank my family who has endured patiently for years as I have pursued my educational goals. My wife, Sandra, and my daughter, Leslie, deserve particular appreciation for their support throughout the process.

Abstract

This research investigates the use of Markov random fields for Bayesian reconstruction algorithms to be used with high-resolution and high-sensitivity SPECT systems for small animal imaging. It extends previous research on mechanical models for Bayesian image reconstruction by using a three-dimensional nonconforming finite element model and linear elasticity concepts to derive minimum potential energy functionals which regularize the reconstruction process. It combines dual collimator SPECT projection data by using high-resolution data to penalize lower-resolution data. It compares the new three-dimensional penalized reconstruction technique with existing penalized techniques through the use of modulation transfer and contrast discrimination functions.

Table of Contents

List of Tables

List of Figures

Nomenclature

Δ	Gradient
$R(f)$	Roughness penalty
Z	Partition function
β	Influence of a prior
T	Temperature
$U(f)$	Gibbs energy function

Chapter 1

Introduction

Single photon emission computed tomography (SPECT) is a non-invasive diagnostic imaging technique for studying physiological processes such as the increased metabolic rate associated with the growth of cancerous tissue or the development of amyloidal plaques within tissue or organs. One advantage of emission tomography is that multiple scans can be taken at different times to study the progression of disease. The goal of this research is to investigate the use of prior data to improve estimation of the location and density of radiopharmaceutical uptake in living tissue, thereby improving diagnostic accuracy. Increased spatial resolution and contrast sensitivity are vital to this estimation process, but a major determinant in spatial resolution is the size of the detector elements used to acquire data and the size of the pinhole collimators. Smaller detector elements capture fewer photons and this results in decreased sensitivity and decreased signal to noise ratios. Reconstruction algorithms have been developed to address these issues by using prior information. Qualitatively superior images are produced by statistical methods using three-dimensional (3-D) data acquisition and reconstruction, and commercial scanners are designed to acquire 3-D data. This increased amount of data results in significant reductions in the image variance; more importantly, it reduces the injected dose of the tracer while maintaining image variance.

Initial tomographic reconstruction was based on convolution or filtered back-projection (FBP), a relatively fast technique but with major disadvantages as it neither models the physics of photon emission and detection, nor does it permit prior knowledge to be incorporated into the reconstruction. It may produce negative intensity values for some voxels. As these negative values have no physical interpretation, they must be truncated to zero, which affects the statistical properties of the resulting image. In addition, FBP requires a large number of equally-spaced view angles so that the frequency space is sampled properly.

Statistical image reconstruction is an attractive alternative to FBP as the physics of photon emission and detection can be modeled accurately. This approach can incorporate anatomical or geometric information, and if it is known that the only possible values of density and attenuation in the body correspond to specific materials, the values can be constrained. Furthermore, it facilitates reconstruction when projection data at different angles is unavailable due to the physical limitations of the scanner geometry. While there may be a lack of detail in portions of the reconstructed image that are undersampled by the projections, there tend to be fewer artifacts spread throughout the image.

Shepp and Vardi (89) introduced a maximum likelihood algorithm based on Expectation Maximization (MLEM). Maximizing the likelihood function in SPECT is equivalent to minimizing the Kullback-Leibler divergence over non-negative vectors and is slow to converge to a solution (14). Bayesian reconstruction techniques permit the use of prior information to influence the image formation process in emission computed tomography. and encourage the reconstructed volume to be represented as a set of piecewise smooth functions. Generalized Expectation Maximization (GEM) (48) combines an image prior with the likelihood function to produce a smooth image. The algorithm computes a maximum *a posteriori* (MAP) estimate or a penalized maximum likelihood estimate.

The use of prior information must be controlled carefully to prevent the image formation process from ignoring data that does not conform to the prior model,

e.g., increased uptake present in a brain scan that indicates the presence of a tumor or decreased uptake in a cardiac scan that represents damaged muscle tissue. The most prevalent use of prior data is to generate images where roughness is penalized thereby giving preference on the assumption that smooth regions are more likely and that uncorrected scatter or detector noise has prevented the creation of smooth regions. One example of a prior model that encourages the formation of smooth regions of interest is the Markov random field, which can be described by the Gibbs probability density function (PDF). The Gibbs prior then becomes a penalty term in the log posterior PDF, which itself is influenced by a weighting factor referred to as a hyperparameter.

MAP reconstruction incorporates the penalty term and defaults to MLEM when the penalty term is a uniform prior. When the penalty term is not convex, the maximization problem does not have a closed form solution; consequently, line search or trust region methods must be used to obtain the ideal solution.

Several researchers have incorporated smooth Gibbs priors into 2-D and 3-D problems (38; 43; 48; 62). Smoothing priors based on the Gibbs distribution are the predominant approach for MAP reconstruction, but anatomical priors have also been used with varying degrees of success. A major shortcoming of the basic Gibbs prior is that it penalizes all abrupt intensity variations as it suppresses noise, which tends to blur image boundaries at the edge of anatomical features. Line processes have been used to control this blurring but their formation can introduce problems if they have been formed using multimodal imaging techniques without registration.

This research investigates the use of prior data based on functional imaging results from different pinhole sizes. By using the projection data acquired using a small pinhole collimator, data acquired using a larger pinhole can be penalized. Penalties based on models from statistical physics are used to smooth and constrain data. The resulting volumetric image reconstructed by a MAP-EM framework provides better spatial resolution and contrast sensitivity than achievable through standard

reconstruction techniques albeit at considerable computational expense. Future efforts will focus upon better optimization techniques and parallel codes.

1.1 Previous research

Statistical image reconstruction has been studied for many years and numerous algorithms have been devised for 3-D PET or SPECT (12; 22; 27; 28; 30; 31; 32; 48; 51; 56; 57; 58; 59; 60; 63; 81; 82; 89). Variational methods based on the *Euler-Lagrange* equations were first explored for use as penalty methods in surface interpolation (8; 45; 92; 93). Lee *et al* (67) presented a mechanism to incorporate prior information derived from co-registered 2-D anatomical X-ray CT or MRI images to improve the image statistics of emission tomography. This research encouraged the reconstructed image to be piecewise smooth by incorporating a penalty term that influenced the smoothing in the vicinity of edges found in the images that represent tissue boundaries. The penalty was modeled as a Gibbs prior distribution with a potential function obtained through a sequence of smooth potential functions. A deterministic annealing algorithm with a closed-form solution was formulated which was the *weak membrane* model from statistical physics and surface interpolation.

Lee (70) built upon this research to create another penalty term for 2-D image reconstruction problems that was based on the *weak plate* model. Line processes were established by using a mean-field annealing approach which formed line processes using a sigmoid transfer function which took directional gradients as input. The potential energy was modeled as the curvature of the plate as proposed by (20; 45). In this model, deformation in the transaxial direction was not considered. One approach to penalizing 3-D reconstructions is to apply the weak plate penalty to each slice of the reconstructed volume, but this approach neglects features in the transaxial direction.

For comparable approaches for penalizing 3-D image reconstructions, the minimum potential energy of a deformed linear elastic body must consider the *Navier-Cauchy* equations which consist of three coupled partial differential equations (PDEs)

Table 1.1: Examples of penalty functions using gradients (Δ)

Convex	
Geman and Geman	$\lvert\Delta\rvert$
Quadratic	Δ^2
Green	$\log\cosh\Delta$
Huber	$\min\{\lvert\Delta\rvert^2, 2\lvert\Delta\rvert - 1\}$
Non-convex	
Geman and McClure	$\frac{\Delta^2}{(1+\Delta^2)}$
Geman and Reynolds	$\frac{\lvert\Delta\rvert}{(1+\Delta)^2}$
Hebert and Leahy	$\log(1+\Delta^2)$
LeClerc	$1+\exp(\Delta/\tau)$
Charbonnier *et al*	$2\sqrt{1+\Delta^2}-2$

that characterize local equilibrium in terms of a displacement vector in curvilinear coordinates. The general solution to these equations has no closed form, but closed-form solutions can be obtained for several specialized problems which have formed the basis for much of the effort in applied solid mechanics.

1.2 Penalty functions from Gibbs priors

A large body of literature exists on the regularization of ill-posed inverse problems and the penalty functions used to perform regularization. These penalty functions use cliques of neighboring pixels or voxels and finite difference schemes to penalize reconstructed images. A partial list is given in Table 1.1.

The seminal research of Geman and Geman (37) introduced the use of line processes. Geman and Reynolds (36) introduced multiple line processes to accommodate multiple dimensions. Intermodality priors have been used by several researchers, but registration and resolution mismatch has limited some results. Anatomical priors identify organs which may not provide the most accurate boundary information for penalized reconstruction.

A high resolution reconstruction penalizes a lower resolution reconstruction through the use of penalties based on Markov random fields. After an initial reconstruction has been performed, a set of line processes is formed and the high resolution data is combined with the lower resolution data to create another reconstruction with better statistics. Other researchers have used anatomical priors to penalize reconstructions. Intramodality priors avoid several of the problems encountered by multimodality penalties.

Often, algorithms for nonlinear optimization problems will compute local, instead of global minimums. Typically, these algorithms will generate a sequence which would converge to such a point if allowed to compute an infinite number of iterations. This research concentrates on gradient-based approaches. Future research might focus on interior-point methods which would be better suited for convex optimization, but perform remarkably well on non-convex optimization problems, too. They usually have local quadratic convergence rates.

For those penalty functions which are nonconvex, special attention must be paid to the algorithms used to optimize the problem. Following the lead of Lee *et al* (67) and Gindi *et al* (39; 40), coordinate descent is used to estimate the line processes and the image. While this approach is provably convergent, it is slow. Another approach employed by Blake and Zisserman (8) is to modify the objective function by approximating a convex penalty function and using nonlinear successive overrelaxation to solve the 2-D optimization problem. Extension of their *Graduated Nonconvexity* (GNC) algorithm to 3-D reconstruction problems poses some problems as the algorithm requires the calculation of the largest eigenvalue for a circulant matrix.

1.3 Penalty functions as energy minimization

Penalty schemes based on mechanical models treat the divergence of reconstructed image from its data as a problem of deformation which releases potential energy,

Table 1.2: Structures in Elasticity Theory

Dimension	Order	
	$m = 1$	$m = 2$
$d = 1$	Strings	Membranes
$d = 2$	Rods	Plates
$d = 3$	Beams	Shells

analogous to the field of applied solid mechanics which seeks to model the deformation of a body (or structure) by elastic forces. Previous research examined the use of the bending energy of plates in the Bayesian reconstruction of emission tomography. These structures and their three-dimensional analogs are given in Table 1.2, but similar approaches cannot be applied in three dimensions. In three dimensions, elasticity problems cannot be solved by a single PDE whereas two-dimensional problems may be solved by a single PDE through the application of the *Airy stress function* to reduce two PDEs to a single one. An alternate approach is to use the 3-D biharmonic equation as a smoothing kernel.

The 1-D linear elasticity problem is modeled as a weak spring for order 1 or weak membrane for order 2. For 2-D, the weak rod or weak plate is considered. In 3-D, weak beam or weak shells are modeled; there are numerous specialized approaches theories in solid mechanics related to the modeling of shells due to the complexity of the linear elasticity problem in three dimensions.

This research is focused on using penalty functions based on mechanical priors or smoothing splines as an approach to penalized SPECT reconstruction. With these penalty functions, it may be possible to combine the high resolution of data acquired from a small diameter collimator, e.g., 0.5 mm, with the high sensitivity of data from a 3.0 mm diameter collimator and develop an algorithm to allow for the simultaneous reconstruction of high-resolution and high-sensitivity data. The steps in such a proposed algorithm are as follows:

Data: System matrices $A_{0.5}$ and $A_{3.0}$, Projection data sets $d_{0.5}$ and $d_{3.0}$
Result: Reconstructed volume V
begin
    Perform dual-head SPECT imaging with different collimators
        Acquire high-resolution data from 0.5 mm collimator
        Acquire high-sensitivity data from 3.0 mm collimator
    Reconstruct high-resolution, low-sensitivity data and extract edge map
    Reconstruct low-resolution, high-sensitivity data using high-resolution edge map
end

Algorithm 1: Dual-head SPECT reconstruction

1.4 Organization

Chapter 2 is an overview of single photon emission computed tomography (SPECT) that describes how the emission process is modeled. Chapter 3 describes the finite element techniques used to model deformation and the biharmonic equations used to smooth data. Chapter 4 describes the use of Markov random fields for Bayesian reconstruction techniques. Chapter 5 discusses the implementation of these techniques to solve the optimization problem. Chapter 6 describes the generation of simulated projection data using the GATE software to model the acquisition of clinical data from a SPECT small animal scanner to test algorithms proposed previously. Quantitative analysis of spatial resolution and contrast sensitivity are presented for some 3-D reconstruction algorithms. Lastly, Chapter 7 gives the conclusions and suggestions for future research.

Chapter 2

An Overview of Single Photon Emission Computed Tomography

Single photon emission computed tomography (SPECT) is a medical imaging modality that uses radionuclides ingested or inhaled by patients. It uses the mathematical theories of tomography and the nuclear properties of radioisotopes to study physiological processes. It is based on the principles of mechanical collimation, as opposed to the electronic collimation techniques employed in Positron Emission Tomography (PET). One advantage of SPECT imaging over PET is that it can potentially allow for simultaneous imaging of multiple radionuclides, since the gamma rays emitted from different radioisotopes can be differentiated based on energy. SPECT has been widely used and validated extensively for myocardial perfusion imaging to diagnose obstructive coronary artery disease. More recently it has been used to identify amyloid plaques in the brain and other organs. This research into the use of Markov random fields to improve the quality of SPECT imaging was motivated by the work of Wall *et al* (97), who have identified radiotracers that bind specifically to hepatic and splenic amyloids in mice based on reconstructed microSPECT/CT images, as well as autoradiographic, isotope biodistribution, and quantitative histochemical analyses. The ability to discern radiographically the

extent of amyloid burden in the mice provides a unique opportunity to evaluate the therapeutic efficacy of pharmacologic compounds designed to inhibit fibril formation or effect amyloid resolution.

2.1 The Physics of SPECT

SPECT uses one or more collimators to acquire data from radionuclides that emit single gamma rays or emit multiple gamma rays with no angular correlations. Budinger and Gullberg (13) were the first to attempt quantitative SPECT by rotating a patient in front of a stationary scintillation camera and applying reconstruction algorithms. Jaszczak *et al* (54) were among the first researchers to conduct SPECT studies with a gamma camera rotating around a patient. Today, a single or dual-head gamma camera, mounted on a rotating gantry, acquires two-dimensional (2D) projection images at equally-spaced angular intervals around a living subject.

The sensitivity of a SPECT system can be improved by incorporating multiple detector heads into the system. Such a system allows multiple angular projections to be acquired simultaneously. In the same acquisition time, each projection can be recorded multiple times, leading to a dramatic increase in the total number of counts detected. Modern SPECT systems have sophisticated gantries that allow the detector heads to follow elliptical orbits or orbits that trace the contour of a living subject rather than a circular orbit. Such orbits allow detector heads to pass closer to the body than would be the case with circular orbits, which can result in significant improvements in sensitivity. In the case of SPECT for small animal imaging, spatial resolution is limited currently to one millimeter whereas modalities such as X-ray CT or magnetic resonance imaging (MRI) have a spatial resolution of around 100 microns (34; 52; 76; 88; 98).

SPECT imaging requires that a patient ingest or inhale a radiopharmaceutical. A scan is begun after sufficient delay to allow the radioisotope to be transported to the tissue or organs of interest. Photon emission is modeled as a spatial Poisson point

process in a pixelated region. For M detectors, d_i, $i \leq M$ is assumed to be a set of independent Poisson variables with means $\overline{d}_i$. The likelihood that photons emitted in f are detected by a detector d_i and counted to produce d is given by

$$P(d|f) = \prod_{i=1}^{M} \exp(-\overline{d}_i) \frac{\overline{d}_i}{d_i!}. \tag{2.1}$$

A linear model is often used to represent the Poisson process

$$d_i \sim \textbf{Poisson}(Af + \eta) \tag{2.2}$$

where d_i is the observed detector counts and η is an estimate of the Gaussian and Poisson noise processes. The system matrix A incorporates the scanner geometry, detector efficiencies and attenuation correction factors (ACFs). The crystal blocks and photon multipliers are not perfect detectors as they are limited by geometric sensitivity problems, crystal penetration and intercrystal scattering. Numerous researchers have modeled detector efficiency to produce a more accurate simulation of the photon detection process in commercially available scanners. In many cases, the scanner geometry is modeled as a large sparse matrix and other factors, such as detector efficiency or ACFs, are modeled as diagonal matrices which are integrated into the image reconstruction algorithms. For this research, the system matrix was created using the approach described in Gregor *et al* (44). Another approach to system matrix creation is to use GATE (53) to generate the coefficients. The individual coefficients of the system matrix can be validated by comparing the projection data generated by projecting the data through the system matrix with projection data generated using Monte Carlo techniques, such as those used by GATE.

Ideally, the signal level for a voxel in a SPECT image would be linearly proportional to the amount of activity contained in the volume of tissue of the subject that corresponds to the location of the voxel. In practice, this result is often not achieved as it is based on several assumptions. First, it is assumed that the line of

response for a collimator hole is an extended cylinder, but the actual response for a collimator hole behaves like a divergent cone. Furthermore, it is assumed that the signal acquired is proportional to the total activity, and the signal from the activity closest to the detector is given more weight than that from deeper tissues due to attenuation. Finally, it is assumed that the activity outside of a line of response does not contribute to the signal for that projected element, but scattered radiation may cause crosstalk between elements.

Some of the discrepancies between idealized assumptions and actual situations in SPECT distort the desired linear relationship between signal level and the amount of activity present. They can also produce artifacts and seriously degrade image quality. Possible solutions include modified approaches to data collection, postprocessing, or the use of iterative reconstruction algorithms which incorporate these factors into the system matrix.

2.2 Transmission and Emission Tomography

Tomography is the reconstruction of a function from its line or plane integrals. In the former, the n-dimensional *Radon transform* $\mathbf{R}$ maps a function on $\mathbb{R}^n$ into the set of its integrals over the hyperplanes of $\mathbb{R}^n$; in the latter, the n-dimensional *X-ray transform* $\mathbf{P}$ maps a function on $\mathbb{R}^n$ into the set of its line integrals (83).

Imaging with penetrating radiation uses a model based on the exponential attenuation law. In transmission tomography, a beam L has an initial intensity I_0 and an intensity I_1 after it has traveled through an attenuating medium, e.g., bone, tissue or air. The function $f(x)$ is the X-ray coefficient of the tissue at point x. The scanning process provides the line integral of the function f along each of the lines L. From these integrals, f can be computed to reconstruct an image of the original volume. The equation can be expressed as

$$I_1 = I_0 \, e^{-\int_L f(x)dx} \tag{2.3}$$

In emission tomography, the radiation sources are inside the volume and the distribution of radionuclides is estimated from the signal detected by external detectors. If μ is the attenuation of the volume, the intensity I outside of the volume is measured by a detector which is collimated so as to detect radiation emitted along the line L and is given by

$$I = e^{-\int_L \mu(y)dy} \int_l f(x)dx \tag{2.4}$$

where $L(x)$ is the section of L between x and the detector.

Radionuclides are designed to target specific physiological processes or tissues to concentrate on regions of interest within a volume.

2.3 The effects of pinhole diameter

System resolution depends upon detector and collimator resolution. It may be calculated as

$$\mathrm{R_{sys}} = \mathrm{R_{coll}} + (b/l)\mathrm{R_{detect}}. \tag{2.5}$$

where b denotes the distance from the pinhole to the center of rotation, l denotes the distance from the pinhole to the detector, and $\mathrm{R_{coll}}$ and $\mathrm{R_{detect}}$ denote collimator and detector resolution, respectively. A single pinhole collimator is the simplest physical structure. The dimensions and location of a pinhole aperture between the source and the detectors play a critical role in the performance of the SPECT imaging system. Collimator performance is affected by the geometry of the pinhole. The preferred pinhole shape is circular or hexagonal to maximize the exposed area of the detector surface for a particular septal thickness. A schematic of single pinhole collimation is provided by Figure 2.1.

Collimator resolution depends upon the size of the effective pinhole aperture and the distance to object. Sensitivity is in proportion to the square of the effective pinhole aperture. Collimator resolution R_{coll} is defined as the full width at half maximum

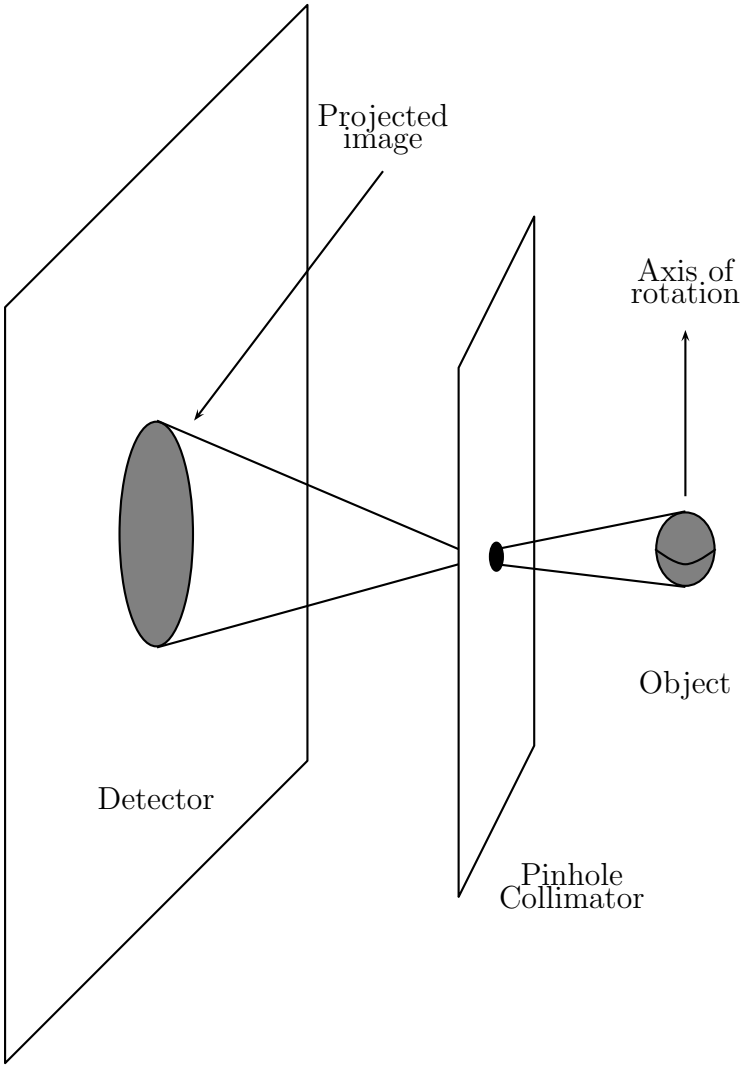


Figure 2.1: A Schematic of Single Pinhole Collimation

(FWHM) of the radiation profile from a point or line source of radiation projected by the collimator onto the detector (16), and it is calculated as

$$\mathrm{R}_{\mathrm{coll}} = d_{\mathrm{eff}}(l+b)/l. \tag{2.6}$$

The effective diameter of the pinhole is calculated as

$$d_{\mathrm{eff}} = \sqrt{d[d + 2\mu^{-1}\tan(\alpha/2)]} \tag{2.7}$$

where d is the pinhole diameter, μ is the linear attenuation coefficient of the pinhole aperture, and α is the cone angle of the pinhole collimator.

The collimator efficiency, ε, is defined as the fraction of gamma rays emitted by the source in the direction of the collimator that passes through it, and it is calculated as

$$\begin{aligned} \varepsilon &= \frac{\pi}{4} d_{\mathrm{eff}} \cos^3\theta / 4\pi b^2. & (2.8)\\ &= d_{\mathrm{eff}} \cos^3\theta / 16 b^2. & (2.9) \end{aligned}$$

Efficiency can be improved by increasing the size of the pinhole, but this results in increased system blur, which iterative reconstruction algorithms using measured point spread functions can compensate partially for pinhole blur.

Resolution improves as the ratio of pinhole diameter to effective length is made smaller; collimator efficiency decreases approximately as the as the square of the ratio of pinhole diameter to length. For a given septal thickness, collimator resolution is improved only at the expense of decreased collimator efficiency; conversely, collimator resolution is decreased as collimator efficiency increases (16). This research attempts to determine whether the tradeoff between system resolution and detector sensitivity can be lessened through the use of penalized reconstruction techniques. As modern preclinical SPECT systems detect a very small fraction of emitted gamma rays, often

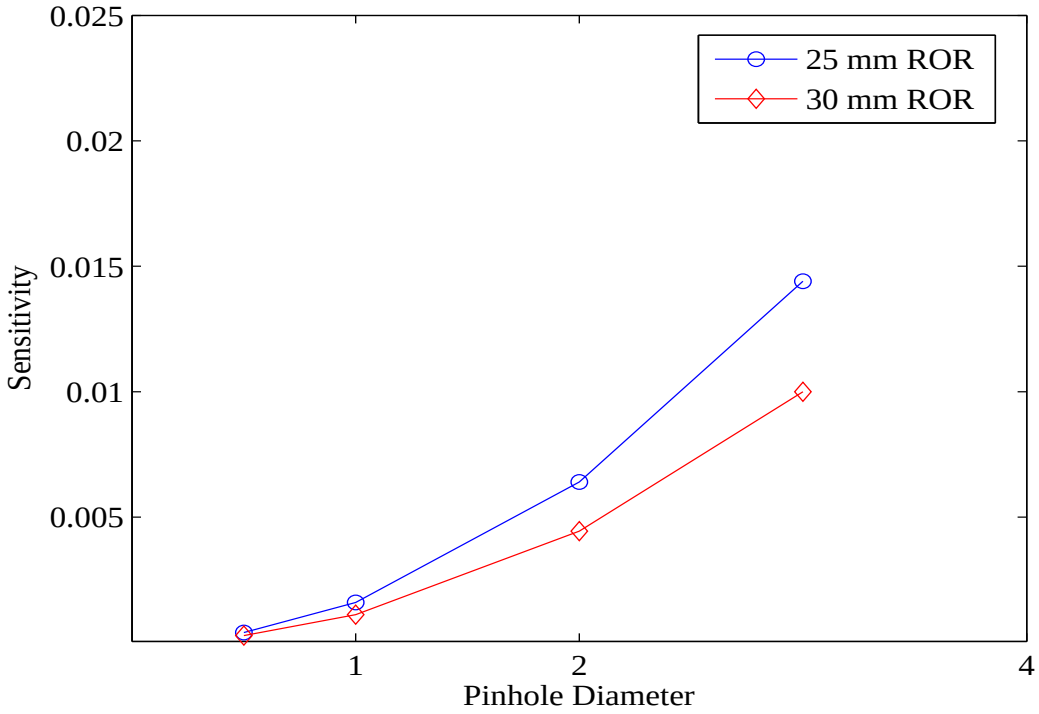


Figure 2.2: Sensitivity vs. Pinhole Diameter for Radius of Rotation

less than 0.1%, this issue is an important one. Figure 2.2 shows the sensitivity of a scanner based on pinhole diameter and the radius of rotation.

2.4 Preclinical Scanners for Small Animal Imaging

The Siemens Inveon Multimodality scanner supports the combination of X-ray, micro-CT, micro-SPECT and PET modalities on a single gantry (41). The micro-SPECT detectors are 10 mm thick pixellated Thallium-doped Sodium Iodide crystals on a 2.2 mm pixel pitch coupled to an array of 2 inch square position-sensitive photomultiplier tubes. These detectors are mounted to the same rotating plane as the CT detectors and both CT and SPECT fields of view are coaxial and directly overlap. As many as four SPECT heads can be placed on the rotating gantry with each detector head mounted to a linear stage. The SPECT module can detect gamma rays from 25

keV to 300 keV, providing sensitivity to most widely used research and clinical single photon isotopes. Key features of the detector head include a large active area (150 mm x 150 mm), which permits greater pinhole magnification, improving sensitivity while maintaining a large field of view; and small detector crystals which provide high intrinsic detector resolution.

The SPECT module has a dual head detector geometry and can be mounted on a rotating gantry. Each detector head contains a 68×68 pixelated scintillator array of 2 mm $\times$ 2 mm pixels. Each crystal is 10 mm thick, providing high sensitivity even to energetic photons. Various interchangeable Tungsten collimators can be attached to the detector. Single pinhole collimators with apertures of 0.5, 1.0, 2.0, and 3.0 mm. The acceptance angle of the single pinhole collimators is 90 degrees, with a focal length of 90 or 95 mm. The transaxial and axial fields of view vary with the radius of rotation from 28 to 45 mm. There are also customized collimators for specific anatomical imaging, e.g.. mouse brain and mouse whole body collimators.

2.5 A Linear Imaging Model

As shown in Equation 2.2, the reconstruction problem can be viewed as a linear imaging model

$$Af + \eta = d \tag{2.10}$$

where A denotes a system matrix which models geometric factors and detector efficiencies, and η which denotes additive noise components such as the Poisson noise from the radioisotope decay process and Gaussian noise produced by electronic components within detectors. The data d represents acquired projection data and f is the unknown tracer concentration which produces a tomographic image. As solving for f is an ill-posed inverse problem, an estimate f^* can be obtained through either least squares or maximum likelihood methods.

The least squares formulation is

$$f^* = \arg\min |Af - d|^2 \tag{2.11}$$

Regularization techniques yield the formulation

$$f^* = \arg\min |Af - d|^2 + \beta R(f) \tag{2.12}$$

where $R(f)$ is a *roughness* penalty term.

The maximum likelihood formulation is

$$f^* = \arg\max \log P(d \mid f) \tag{2.13}$$

Penalized techniques yield the formulation:

$$f^* = \arg\max \log P(d \mid f) + \beta \log P(f). \tag{2.14}$$

where $P(f) \equiv \exp\{-R(f)\}$. The parameter β is typically a user-defined hyperparameter that signifies the influence of the penalty term; it may be estimated using various techniques.

Chapter 3

Finite Element Methods

There is a large body of research in image analysis that attempts to develop physics-based models to describe and interpret contours and surfaces in 2-D and 3-D imagery (77). Deformable models have been used extensively in image segmentation, tracking and registration (87). These models are motivated by the physical phenomena of solid objects being deformed under applied loads. In some applications, the deformation models an object's change of shape over a period of time, i.e., tracking; in others, the difference between two similar objects, i.e., registration.

In surface reconstruction, a series of images taken at different angles are combined to create a 3-D representation of the scene. Grimson (45) and Terzopoulos (92; 93) characterized the problem of surface reconstruction in stereo vision and proposed solutions based on spline interpolation in one and two dimensions. Noting the similarity to physical interpretations of elasticity theory, they sought to model surfaces by calculating the deflection of a thin plate at equilibrium as the minimization of the total potential energy of the bending plate.

3.1 Conforming and Nonconforming Finite Elements

Finite element methods have been developed to solve variational problems such as the deformation of an elastic body. The *Rayleigh-Ritz method* has been used to approximate the solution to boundary value problems for elliptic partial differential equations. In this method, a finite dimensional subspace S is chosen such that

$$S = S^h = \{v^h | v^h = \sum_{i=1}^{N} v_i \phi_i\} \tag{3.1}$$

where the functions $v^h \in S^h$ are piecewise polynomials over S^h, $\phi_1, \ldots, \phi_N$ are independent basis functions which span S and $v_1, \ldots, v_N$ are unknown real parameters. Terzopoulos (92) notes that the basis functions ϕ_i are often complicated functions with global support over the domain while the finite element method can use simple basis functions with local support.

In the development of the finite element method for a second-order differential equation, piecewise polynomials in a finite element space V_h are required to be continuous through the domain Ω. Due to this continuity requirement, the resulting method is called the H^1-conforming finite element method. For the discretization of a fourth-order problem, such as the biharmonic equation, functions in V_h and their first derivatives are required to be continuous on $\bar{\Omega}$. In this case, the finite element method is called the H^2-conforming method. For those cases where the functions in V_h are not required to be continuous on $\bar{\Omega}$ for the discretization of a second-order problem, a *nonconforming* finite element method may be used.

Following the example of Terzopoulos (92), consider an arbitrary patch of eight adjacent elements which share a common vertex $v_{i,j,k}$ internal to the patch, as shown in Figure 3.2. Suppose that a constant strain condition is applied to the patch, i.e., suppose that the displacements at all remaining vertices around the periphery

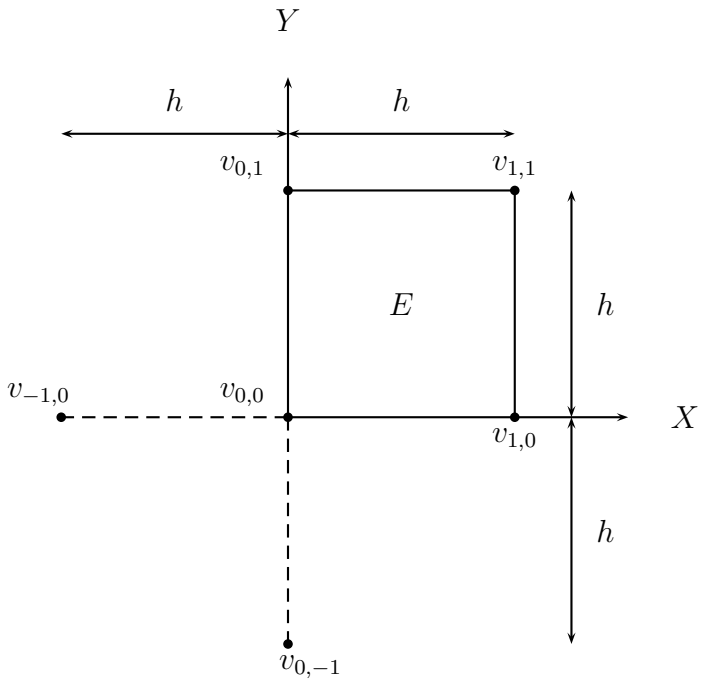


Figure 3.1: Two-dimensional nonconforming finite element

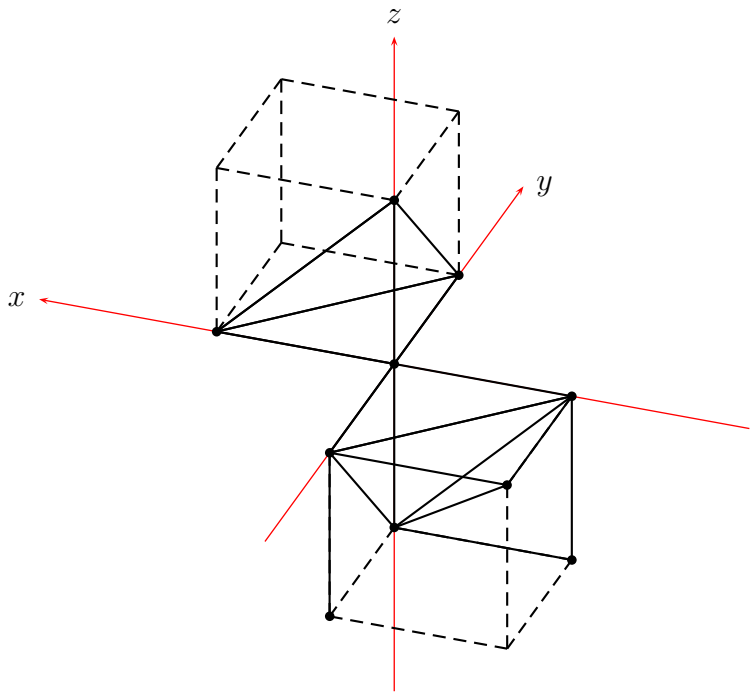


Figure 3.2: A 3-D Nonconforming Finite Element

of the patch are constrained by assigning them values consistent with an arbitrary second-degree polynomial π_2.

Definition 1. *Following the definition in (92), the three-dimensional nonconforming finite element is defined as follows:*

Given that $\bar{\Omega}$ *is rectangular, consider a* uniform *triangulation* $\mathcal{T}^h$ *of* $\bar{\Omega}$ *into identical square elements* E*, where the fundamental length* h *is the length of each side of* E*. By definition,* $\bigcup_{E\in\mathcal{T}^h} E = \bar{\Omega}$ *and the elements are adjacent and overlap along their sides. A point in* $\bar{\Omega}$ *is a node of the triangulation if it is a vertex of an elemental square, and the elements are considered to be interconnected at the nodes. The nodal variables are the* node displacements*, i.e., the values of the function* $v^h \in S^h$ *at the nodes.*

Terzopoulos (92) defines a space P^E of polynomials p^E over the domain Ω which must satisfy the completeness condition in Section which states that $\Pi^2 \in P^E$ since the energy inner product contains derivatives of order $m = 2$. This is the requirement

that all of the polynomials must be able to replicate all states of constant strain, which are all of the polynomials up to degree 2. This requirement is satisfied by choosing p^E to be the six-dimensional space of full second degree polynomials $p^E : E \mapsto \Re$. A two-dimensional polynomial transformation p^E has six parameters from a to f which Terzopoulos used to determine the node displacements for a plate. It is represented as follows:

$$p^E(x, y) = ax^2 + by^2 + cxy + dx + ey + f. \tag{3.2}$$

Terzopolous solves a system of equations to arrive at the values

$$\begin{aligned}
a &= \frac{1}{2h^2}(v_{1,0} - 2v_{0,0} + v_{-1,0}) && (3.3)\\
b &= \frac{1}{2h^2}(v_{0,1} - 2v_{0,0} + v_{0,-1}) && (3.4)\\
c &= \frac{1}{h^2}(v_{1,1} - v_{0,1} - v_{1,0} + v_{0,0}) && (3.5)\\
d &= \frac{1}{2h}(v_{1,0} - v_{-1,0}) && (3.6)\\
e &= \frac{1}{2h}(v_{0,1} - v_{0,-1}) && (3.7)\\
f &= v_{0,0} && (3.8)
\end{aligned}$$

The three-dimensional polynomial transformation takes the following form with ten parameters from a to j:

$$p^E(x, y, z) = ax^2 + by^2 + cz^2 + dxy + exz + fyz + gx + hy + iz + j. \tag{3.9}$$

In a similar manner, a system of ten equations can be solved to arrive at the finite difference formulas in Equations 3.10 to 3.19. The vertex $p_{1,1,1}$ is a linear combination

of the other equations and may be omitted from consideration.

$$a = \frac{1}{2h^2}(v_{1,0,0} - 2v_{0,0,0} + v_{-1,0,0}) \tag{3.10}$$

$$b = \frac{1}{2h^2}(v_{0,1,0} - 2v_{0,0,0} + v_{0,-1,0}) \tag{3.11}$$

$$c = \frac{1}{2h^2}(v_{0,0,1} - 2v_{0,0,0} + v_{0,0,-1}) \tag{3.12}$$

$$d = \frac{1}{h^2}(v_{1,1,0} - v_{0,1,0} - v_{1,0,0} + v_{0,0,0}) \tag{3.13}$$

$$e = \frac{1}{h^2}(v_{1,0,1} - v_{0,0,1} - v_{1,0,0} + v_{0,0,0}) \tag{3.14}$$

$$f = \frac{1}{h^2}(v_{0,1,1} - v_{0,0,1} - v_{0,1,0} + v_{0,0,0}) \tag{3.15}$$

$$g = \frac{1}{2h}(v_{1,0,0} - v_{-1,0,0}) \tag{3.16}$$

$$h = \frac{1}{2h}(v_{0,1,0} - v_{0,-1,0}) \tag{3.17}$$

$$i = \frac{1}{2h}(v_{0,0,1} - v_{0,0,-1}) \tag{3.18}$$

$$j = v_{0,0,0} \tag{3.19}$$

Terzopoulos (93) states that the continuous surface reconstruction problem may be approximated by a discrete variational approach since a closed-form solution is infeasible due to the irregular occurrence of constraints and discontinuities. His research employs the finite element method as a local approximation technique. Terzopoulos (93) later introduced a class of multidimensional *controlled-continuity stabilizers*, consisting of generalized spline kernels for the regularization of inverse visual problems which are based upon the research of Duchon (25; 26) and Meinguet (78).

The following theorems and lemma from Ciarlet (18) form the basis for the use of both conforming and nonconforming finite elements.

Theorem 3.1 (Existence and uniqueness of conforming finite elements)**.** *If there exists a mapping energy inner products are uniformly S^h-elliptic, there exists a*

constant C, independent of S^h such that

$$||u - u^h|| \leq C \left(\inf v^h \in S^h ||u - v^h|| + \sup w^h \in S^h \frac{|a_h(u, v^h) - f(w^h)|}{||w^h||} \right)^{\frac{1}{2}}. \quad (3.20)$$

Proof. See (18, p. 210). □

Theorem 3.2 (Lax-Milgram lemma)**.** *Let V be a Hilbert space, let $a(\cdot, \cdot) : V \times V \to R$ be a continuous, V-elliptic bilinear form, and let $f : V \to R$ be a continuous linear form. Then the abstract variational form may be stated as:*

Find an element u such that $u \in V$ and $\forall v \in V, a(u, v) = f(v)$, has one and only one solution.

Proof. See (18, pp. 8-9). □

Lemma 3.2.1 (Céa's lemma)**.** *There exists a constant C independent upon the subspace S^h such that*

$$||u - u^h|| \leq C \inf_{v^h \in S^h} ||u - v^h||. \quad (3.21)$$

Consequently, a sufficient condition for convergence is that there exists a family (S^h) of subspaces of the space S such that, for each $u \in S$,

$$\lim_{h \to 0} \inf_{v^h \in S^h} ||u - v^h|| = 0. \quad (3.22)$$

Proof. See (18, pp. 104-105). □

Theorem 3.1 is extended to nonconforming elements by Theorem 3.3 (18).

Theorem 3.3 (second Strang lemma)**.** *Consider a family of discrete problems for which the associated approximate bilinear forms are uniformly S^h-elliptic.*

Then, there exists a constant C, independent of S^h such that

$$||u - u^h||_h \leq C \left(\inf_{v^h \in S^h} ||u - v^h||_h + \sup_{w^h \in S^h} \frac{|a_h(u, w^h) - f(w^h)|}{||w^h||_h} \right). \quad (3.23)$$

Proof. See (18, p. 210). □

Proposition 1. *The nonconforming finite element of Definition 1 converges.*

No proof of Proposition 1 is given; however, the finite difference formulas in Equations 3.10 to 3.19 can be derived by Taylor series expansions of the second-order partial derivatives.

3.2 Thin-plate Interpolating Splines and the Biharmonic Equation

The thin-plate interpolating spline $J_m^d(u)$ that models the deformation has a unique minimal solution under the condition that the least-squares regression on $\phi_1, \ldots, \phi_M$ is unique (25; 26; 96). In two or three dimensions, this dictates that all points may not lie on the same line or plane, respectively. The solution can be stated analytically as

$$u(x) = \sum_{\nu=1}^{M} a_\nu \phi_\nu(x) + \sum_{i=1}^{M} w_i U(x, p_i) \tag{3.24}$$

with basis function $U(x, p_i)$ depending on m, d and the Hilbert space $\mathcal{H}$ of admissible functions.

Image registration techniques use elastic transformations based on thin-plate splines. Thin-plate splines have a physical interpretation since they minimize the bending energy of a thin plate. These splines serve as a flexible deformation model which yield a transformation of $\mathcal{C}^1$-continuity with minimal curvature. The mathematical theory of thin-plate splines is well understood. Thin plate splines result as a function of an optimization problem which is well posed, i.e., the solution exists, is unique and depends continuously upon the data. Moreover, the solution can be stated in closed form. These splines even have a statistical interpretation as they are Bayesian estimates with a prior (96).

This research uses thin-plate splines, but other researchers use radial basis functions to calculate the displacement fields. These functions span an n-dimensional space of functions that depend upon the vertices $\mathbf{v}$ of the images. Choosing the space of functions on $\mathbb{R}^d$ for which all possible derivatives of total order m are integratable, results in basis functions

$$U(r) = \begin{cases} \theta_{m,d} r^{2m-d} \ln r & r \text{ even} \\ \theta_{m,d} r^{2m-d} & r \text{ odd} \end{cases} \tag{3.25}$$

where

$$\theta_{m,d} = \begin{cases} \dfrac{(-1)^{d/2+1+m}}{2^{2m-1}\pi^{d/2}(m-1)!(m-d/2)!} \\ \dfrac{\Gamma(d/2-m)}{2^{2m}\pi^{d/2}(m-1)!} \end{cases} \tag{3.26}$$

Γ denotes the Gamma function and r

$$r = |x - v| = \sqrt{\sum_{k=1}^{d} (x_k - v_k)^2} \tag{3.27}$$

yields the identity

$$U(x, v) = U(|x - v|) = U(r) \tag{3.28}$$

The function $U(r)$ is a *Green's function* for the m-th iterated Laplacian and solves the equation

$$\Delta^m U(r) = \delta(r), \tag{3.29}$$

where $\delta(r)$ is the Dirac delta function, which is the fundamental solution to the biharmonic equation. If the domain is bi-infinite, then it is computed as (8, p. 184).

$$G(x, \acute{x}) = \frac{1}{2\lambda} \exp(-|x - \acute{x}|/\lambda). \tag{3.30}$$

Let $f : \Omega \mapsto \mathbb{R}$ be a smooth function on the region $\Omega \in \mathbb{R}^3$. Let $f_0 = (x, y, z, t) \in \Omega$ at time t and $f_1 = (x + u, y + v, z + w, t + \delta t)$ be f_0 which has been deformed by the displacement vector $\mathbf{u} = (u, v, w)$ through some force $\mathbf{f}$.

Terzopoulos formulated multivariate smoothness constraints based on spline approximations of the following class of functionals originally proposed by Duchon (25; 26) and Meinguet (78). For $m = 1, d = 2$, the functional reduces to

$$|v|_1^2 = \int \int_{R^2} (v_x^2 + v_y^2)\, dx\, dy, \tag{3.31}$$

which is proportional to the small deflection energy of a membrane. The associated Euler-Lagrange equation is *Laplace's equation*,

$$\nabla u = 0, u = \frac{\partial^2 f}{dx^2} + \frac{\partial^2 f}{dy^2} \tag{3.32}$$

For $m = 2, d = 2$, the functional reduces to

$$|v|_2^2 = \int \int_{R^2} (v_x^2 + 2v_{xy}^2 + v_y^2)\, dx\, dy, \tag{3.33}$$

which is proportional to the small deflection bending energy of a thin plate with zero Poisson ratio. The associated Euler-Lagrange equation is the *biharmonic equation*,

$$\nabla^4 u = (\nabla^2)^2 = 0, u = \frac{\partial^4 f}{dx^4} + \frac{\partial^4 f}{dx^4 dy^4} + \frac{\partial^4 f}{dy^4} \tag{3.34}$$

Duchon (25; 26) refers to the minimizers of $|v|_2^2$ as "thin-plate splines."

For $m = 2, d = 3$, the functional reduces to

$$|v|_3^2 = \int \int \int_{R^3} \left(v_x^2 + v_y^2 + v_z^2 + 2[v_{xy}^2 + v_{xz}^2 + v_{yz}^2]\right)\, dx\, dy\, dz. \tag{3.35}$$

The associated Euler-Lagrange equation is the 3-D biharmonic equation,

$$\nabla^4 u = (\nabla^2)^2 = 0, u = \frac{\partial^4 f}{dx^4} + \frac{\partial^4 f}{dy^4} + \frac{\partial^4 f}{dz^4} + \frac{\partial^4 f}{dx^4 dy^4} + \frac{\partial^4 f}{dx^4 dz^4} + \frac{\partial^4 f}{dy^4 dz^4}. \quad (3.36)$$

Bookstein (9) refers to the minimizers of $|v|_3^2$ as "thin-hyperplate splines."

Equation 3.34 has Dirichlet boundary conditions of the first kind

$$u = g_1(x, y), \frac{\partial u}{\partial n} = g_2(x, y), (x, y) \in \Omega \quad (3.37)$$

Here Ω is a closed convex domain in two dimensions and $\partial\Omega$ is its boundary. Similarly, Equation 3.36 has Dirichlet boundary conditions of the first kind

$$u = g_1(x, y, z), \frac{\partial u}{\partial n} = g_2(x, y, z), (x, y, z) \in \Omega \quad (3.38)$$

Here Ω is a closed convex domain in three dimensions and $\partial\Omega$ is its boundary.

Equation 3.34 can be approximated by a thirteen-point finite difference stencil on a 5×5 grid with modifications at the boundaries (1, p. 885). This approximation has been used to solve the surface reconstruction problem in two dimensions (45; 92; 93). Several researchers have commented on the difficulty in constructing continuously-differentiable finite elements in two and three dimensions. For the biharmonic equation, nonconforming finite elements to lessen computational requirements. Lascaux and Lesaint (64) described a number of nonconforming finite element methods for the biharmonic equation in $\mathbf{R}^2$.

Chapter 4

Markov Random Fields

Markov random fields (MRFs) have been used in image restoration and segmentation, edge detection, texture analysis, stereo matching (90), surface reconstruction (35) and optical flow computation (61; 50). MRFs can model the *a priori* probability of contextual-dependent patterns such as textures. Maximum a posterior (MAP) is one of the most popular statistical criteria for optimality and has been the most popular choice for MRF vision modeling. The MAP-MRF framework advocated by Geman and Geman (37) models the objective as the joint posterior probability of the MRF labels. MAP-MRF modeling derives the form of the posterior distribution and determines its parameters in order to define the posterior probability. The maximum value of the posterior distribution is then obtained using optimization techniques or stochastic techniques.

4.1 Preliminaries

Let $F = \{F_1, \ldots, F_m\}$ be a random field defined on a set S where each random variable F_i takes a value in a set of labels $\mathcal{L}$. Following the notation of (71), $F_i = f_i$ denotes the random variable F_i taking the variable f_i and the joint event $(F_1 = f_1, \ldots, F_m = f_m)$, abbreviated as $F = f$ where $f = \{f_1, \ldots, f_m\}$, represents a *configuration* of F, corresponding to a realization of the field. For a discrete label set $\mathcal{L}$, the probability

that a random variable F_i takes the value f_i is denoted as $P(F_i = f_i)$ and abbreviated as $P(f_i)$. The joint probability is denoted as $P(F = f) = P(F_1 = f_1, \ldots, F_m = f_m)$ and abbreviated as $P(f)$. For the corresponding continuous label sets, probability density functions (p.d.f.'s) are denoted $p(F_i = f_i)$ and $p(F = f)$, respectively. Some configurations have a greater likelihood of occurrence than others and some configurations may be prohibited.

A Markov random field is a stochastic process that obeys a conditional relation given by

$$\begin{aligned} & P\left(F_{i,j,k} = f_{i,j,k} \mid \{F_{l,m,n} = f_{l,m,n} \mid (l,m,n) \neq (i,j,k)\}\right) \\ & = P\left(F_{i,j,k} = f_{i,j,k} \mid \{F_{l,m,n} = f_{l,m,n} \mid (l,m,n) \in \mathcal{N}_{i,j,k}\}\right), \end{aligned} \tag{4.1}$$

where $\mathcal{N}_{i,j,k}$ denotes the set of neighbors of the location (i,j,k). The site (l,m,n) where $l \neq i, m \neq j, n \neq k$ is defined as a neighbor of site (i,j,k) if and only if the functional form of

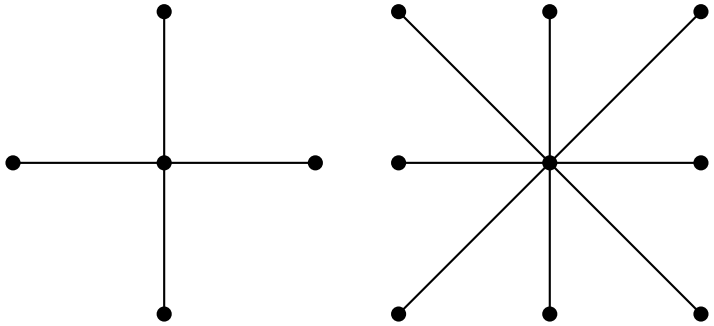
$$P\left(F_{i,j,k} = f_{i,j,k} \mid \{F_{l,m,n} = f_{l,m,n} \mid (l,m,n) \neq (i,j,k)\}\right) \tag{4.2}$$

depends upon $f_{l,m,n}$ (4). The conditional distribution of $F_{i,j,k}$ is locally dependent and symmetry must be preserved among neighbors. A set of sites that satisfies these conditions is called a *clique* $\mathcal{C}$. A unary clique is denoted $\mathcal{C}_1$, a pairware clique is denoted $\mathcal{C}_2$ and higher-order cliques are denoted $\mathcal{C}_3$, $\mathcal{C}_4$,$\ldots$,$\mathcal{C}_n$.

Geman and Geman (37) define neighborhoods $\mathcal{N}_k$ as zeroth order ($\mathcal{N}_0$), first order ($\mathcal{N}_1$), second order ($\mathcal{N}_2$), $\ldots$, kth order ($\mathcal{N}_k$) depending on the form

$$\mathcal{N}_k = \left\{(l,m,n) \mid 0 < (l-i)^2 + (m-j)^2 + (n-k)^2 \leq k\right\} \tag{4.3}$$

where (l,m,n) is a neighbor of (i,j,k). Figure 4.3 and Figure 4.4 illustrate first, second and third-order MRF neighborhoods corresponding to two and three dimensions.



(a) First order **(b)** Second order

Figure 4.1: 2-D MRF Neighborhoods

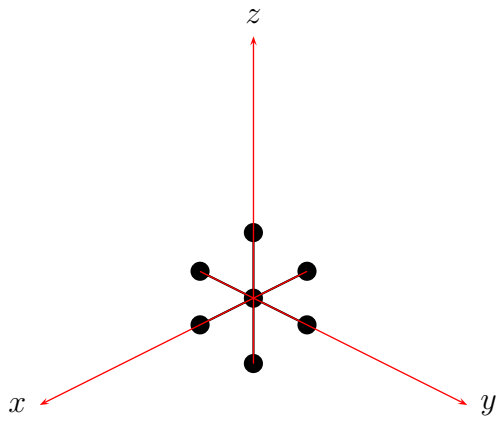


Figure 4.2: First-order 3-D MRF Neighborhood

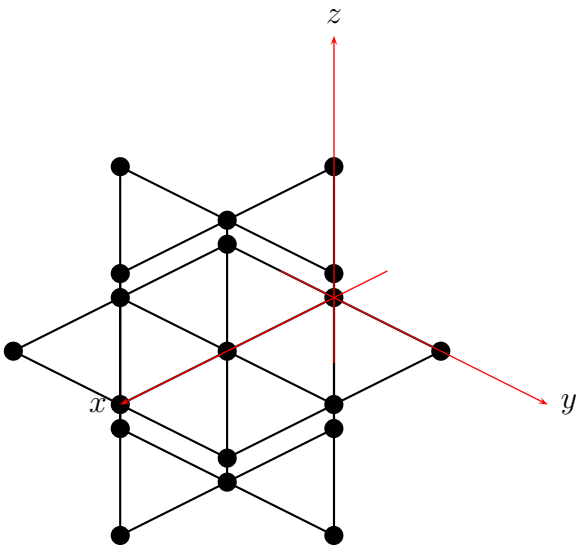


Figure 4.3: Second-order 3-D MRF Neighborhood

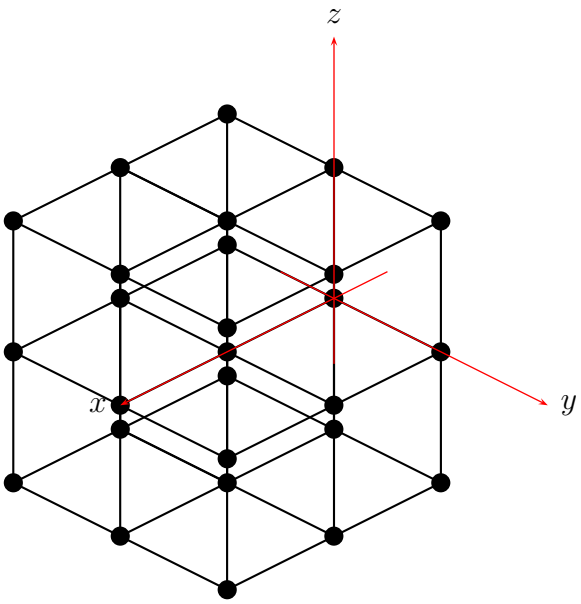


Figure 4.4: Third-order 3-D MRF Neighborhood

Definition 2. *(71, pp. 24-25) F is a Markov random field on S with respect to a neighborhood $\mathcal{N}$ if and only if the following conditions are satisfied:*

1. $P(f) > 0, \forall f \in \mathbb{F}$ *(positivity)*

2. $P(f_i|f_{S-\{i\}}) = P(f_i|f_{\mathcal{N}_i})$ *(Markovianity)*

where $S - \{i\}$ is the set difference, $f_{S-\{i\}}$ denotes the set of labels in $S - \{i\}$ and $f_{\mathcal{N}_i} = \{f_i | i \in \mathcal{N}\}$.

When the positivity condition is satisfied, the joint probability $P(f)$ of any random field is uniquely determined by its local conditional probabilities (4). The Markovianity condition depicts the local characteristics of F and limits interactions to neighboring pixels or voxels.

4.2 Markov Random Fields and the Gibbs Distribution

The prior distribution of a random variable f can be specified as a Gibbs distribution

$$P(f) = \frac{1}{Z} \exp\left\{-\frac{1}{T}U(f)\right\}. \tag{4.4}$$

where Z is a normalizing constant

$$Z = \sum_{f \in \mathbb{F}} \exp\left\{-\frac{1}{T}U(f)\right\}. \tag{4.5}$$

T is a control parameter called a temperature and $U(f)$ is a Gibbs energy function defined on the set $\Omega = \{f(i,j) \in R \mid f(i,j) \geq 0, i = 1, 2, \ldots, N\}$. The energy function $U(f)$ is given by

$$U(f) = \sum_{c \in \mathcal{C}} V_c(f) \tag{4.6}$$

which is the sum of clique potentials $V_c(f)$ over all possible cliques $\mathcal{C}$. The value of $V_c(f)$ depends on the local configuration of the clique c. The partition function Z, which is the sum over all possible configurations in $\mathbb{F}$, must be evaluated to calculate the Gibbs distribution. Since this evaluation is prohibitive, several approximation techniques have been developed to solve this problem (71).

The energy of a Gibbs distribution may be expressed as the sum of several terms, each ascribed to cliques of a certain size:

$$U(f) = \sum_{\{i \in \mathcal{C}_1\}} V_1(f_i) + \sum_{\{i,j\} \in \mathcal{C}_2} V_2(f_i, f_j) + \sum_{\{i,j,k\} \in \mathcal{C}_3} V_3(f_i, f_j, f_k) + \cdots \tag{4.7}$$

When only unary or pairwise cliques are considered, the energy may be written as

$$U(f) = \sum_{i \in S} V_1(f_i) + \sum_{i \in S} \sum_{j \in \mathcal{N}_i} V_2(f_i, f_j) \tag{4.8}$$

In the second term of the right hand side of Equation 4.8, $\{i, j\}$ and $\{j, i\}$ are two distinct cliques in $\mathcal{C}_2$ because the sites in a clique are *ordered*.

The conditional probablity of Equation 4.8 may be written as

$$P(f_i \mid f_{\mathcal{N}_i}) = \frac{\exp - \left\{ V_1(f_i) + \sum_{j \in \mathcal{N}_i} V_2(f_i, f_j) \right\}}{\sum_{f_i \in \mathcal{L}} \exp - \left\{ V_1(f_i) + \sum_{j \in \mathcal{N}_i} V_2(f_i, f_j) \right\}} \tag{4.9}$$

By incorporating Equation 4.8 into Equation 4.4, the joint probability can be written as (71)

$$P(f) = Z^{-1} \prod_{i \in \mathcal{S}} r_i(f_i) \prod_{i \in \mathcal{S}} \prod_{j \in \mathcal{N}_i} r_{i,j}(f_i, f_j) \tag{4.10}$$

where $r_i(f_i) = \exp - \left\{ \frac{1}{T} V_1(f_i) \right\}$ and $r_{i,j}(f_{i,j}) = \exp - \left\{ \frac{1}{T} V_2(f_i, f_j) \right\}$.

A special case of Equation 4.8 is called an *auto-model* (4) when $V_1(f_i) = f_i G_i(f_i)$ and $V_2(f_i, f_j) = \beta(f_i, f_j)$, where $G_i(\cdot)$ are arbitrary functions and $\beta_{i,j}$ are constants reflecting the pairwise interaction between i and j:

$$U(f) = \sum_{\{i\} \in \mathcal{C}_1} f_i G_i(f_i) + \sum_{\{i,j\} \in \mathcal{C}_2} \beta_{i,j}(f_i, f_j) \tag{4.11}$$

An auto-model is considered an *auto-logistic* model if the $f_i's$ take on values in the discrete label set $\mathcal{L} = \{0, 1\}$ or $\mathcal{L} = \{-1, +1\}$ in the case of the *Ising model*. The corresponding energy is of the form

$$U(f) = \sum_{\{i\} \in \mathcal{C}_1} \alpha_i G_i(f_i) + \sum_{\{i,j\} \in \mathcal{C}_2} \beta_{i,j}(f_i, f_j) \tag{4.12}$$

where $\beta_{i,j}$ is said to be the *interaction coefficients* (71). When $\mathcal{N}$ is the nearest neighborhood system on a 1D or 2D lattice, the auto-logistic model is reduced to the *Ising model*. The auto-logistic model can be generalized further to a *multilevel logistic* (MLL) model or *generalized Ising model* (37). There are $M(> 2)$ discrete labels in

the label set $\mathcal{L} = \{1, 2, \ldots, M\}$. In this model, a clique potential depends upon the type of clique and the local configuration. For cliques containing more than one site, the MLL clique potentials are defined as

$$V_c(f) = \begin{cases} \zeta_c & \text{if all sites on } c \text{ have the same label} \\ -\zeta_c & \text{otherwise} \end{cases} \tag{4.13}$$

where ζ_c is the potential for type c cliques; for single-site cliques, they depend on the label assigned to the site

$$V_c(f) = V_c(f_i) = \alpha_I \text{ if } f_i = I \in \mathcal{L}_d \tag{4.14}$$

where α_I is the potential for the label value I.

If the MLL model is of second order, the potential function is written as

$$V_2(f_i, f_j) = \begin{cases} \beta_c & \text{if sites on clique } \{i, j\} = c \in \mathcal{C}_2 \text{ have the same label} \\ -\beta_c & \text{otherwise} \end{cases} \tag{4.15}$$

where β_c is the β parameter for type c cliques and $\mathcal{C}_2$ is the set of pairwise cliques. When the model is isotropic, the neighbors all take the same value. Setting all unary clique potentials to zero, the pairwise clique potentials are used with a potential function $g(\cdot)$ to encode piecewise smoothness that satisfies the necessary condition

$$\lim_{\eta \to \infty} |g\prime(\eta)| = \lim_{\eta \to \infty} |2\eta h(\eta)| = C < \infty \tag{4.16}$$

where $C \in [0, \infty]$ is a constant (71, p. 53). Potential functions such as the truncated quadratic of Blake and Zisserman (8) are described in more detail in Section 4.6.

4.3 Higher-order Cliques

Higher-order cliques have been developed to solve numerous problems, and researchers continue to devise techniques to cope with the increased computational requirements. The multi-level logistic models that form the basis for smoothness priors are based on the generalized Ising model. For some applications, the higher-order cliques must be considered.

Hierarchical two-level Gibbs models have been proposed to represent both textured and noise-contaminated images (23; 24; 73). The higher level Gibbs distribution uses an isotropic random field, e.g. MLL, to characterize the blob-like region formation process. A lower level Gibbs distribution describes the filling-in in each region. The filling-in may be independent noise or a type of texture, both of which can be characterized by Gibbs distributions. This provides a convenient approach for MAP-MRF modeling. In segmentation of noisy and textured images, the higher level often determines the prior of f for the region process while the lower level Gibbs random field contributes to the conditional probability of the data given f. Different levels of MRFs in the hierarchy can have different neighborhood systems.

Spatiotemporal MRF models extend the MRF lattice in three dimensions by adding adjacent GRFs to represent adjacent time intervals (72). Such MRFs might be employed to manage dynamic SPECT imaging.

4.4 Clique Potential Functions

According to the Hammersley-Clifford theorem, specifying the prior distribution of an MRF amounts to specifying the clique potential functions V_C in the corresponding Gibbs distribution in Equation 4.4. The form of the clique potential function depends upon the size of the neighborhood $\mathcal{N}$ and the potential energy function $\phi(\cdot)$ which penalizes the violation of smoothness caused by the difference between labels. Energy functions are typically linear, quadratic, or convex nonquadratic. A truncated linear

or quadratic function as in (37; 8) may be used to allow for spatial discontinuity (e.g., edges).

The typical MRF formulation has a finite difference Δ as the argument to the potential function $\phi(\cdot)$. The first order finite difference yields

$$\begin{aligned} V_c(i,j) &= \sum_{\{i,j\}} (f_i - f_j)^2 \\ &= \sum_{\{i,j\}} \phi(\Delta) \text{ where } \phi(x) = f^2, \Delta = \tfrac{\partial f}{\partial x} \end{aligned} \tag{4.17}$$

In an effort to restore images corrupted by noise, Geman and Geman (37) devised a binary line process for Markov Random Fields (MRFs) which formed a dual lattice along with the neighborhood for the intensity process. The purpose of the binary line model was to prevent the smoothing of sharp discontinuities in intensity values due to edge transitions. The binary line process model was often created by gradient operators. Any penalty function applied to the image would not be applied to any voxel belonging to the line process model. In this case, the first-order finite difference yields

$$\begin{aligned} V_c(i,j) &= \sum_{\{i,j\}} (x_i - x_j)^2 (1 - z_{ij}) \\ &= \sum_{\{i,j\}} \phi(\Delta)(1 - z_{ij}) \text{ where } \phi(x) = x^2, \Delta = \tfrac{\partial f}{\partial x} \end{aligned} \tag{4.18}$$

The optimization problem posed by the mixed-mode clique potentials in Equation 4.18 is difficult to solve, and it was reformulated by Blake and Zisserman (8) to

obtain a convex approximation:

$$\begin{aligned} V_c(i,j) &= \sum_{\{i,j\}} (x_i - x_j)^2 (1 - z_{ij}) \\ &= \sum_{\{i,j\}} g(x_i - x_j, z_{ij}) \\ &= \sum_{\{i,j\}} \min\left\{(x_i - x_j)^2, \alpha\right\} + \alpha \sum_{\{i,j\}} z \\ &= \sum_{\{i,j\}} -\max\left\{-(x_i - x_j)^2, -\alpha\right\} + \alpha \sum_{\{i,j\}} z \end{aligned} \tag{4.19}$$

The function $g(\cdot)$ must be a convex function in order for a closed-form solution to exist for the optimization problem. The *log-sum-exp* function $\log\exp(\cdot) + \cdots + \exp(\cdot))$ is a convex function and can be interpreted as a differentiable approximation of the $\max(\cdot)$ function (10). Substituting the Laplacian operator for the finite difference results in the biharmonic operator serving as the input to $\log\exp(\cdot) + \cdots + \exp(\cdot))$:

$$\begin{aligned} V_c(i,j) &= \sum_{\{i,j\}} g(\Delta) \\ &= \sum_{\{i,j\}} \log\{\exp(\nabla^2)^2 + \exp(\alpha)\} \\ &= \sum_{\{i,j\}} \log\{\exp(\nabla^4) + \exp(\alpha)\} \end{aligned} \tag{4.20}$$

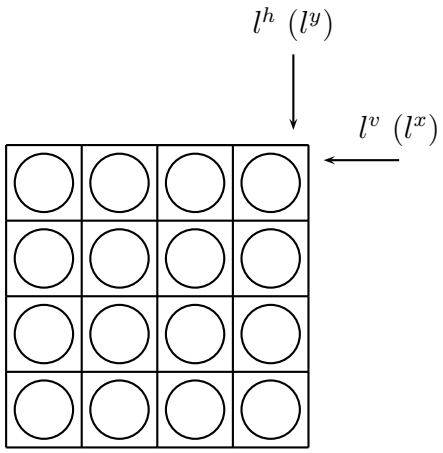


Figure 4.5: 2-D Line Processes

Since $g(u) = \exp\{-\beta u\}$ is a convex function, *Jensen's inequality* holds,

$$\begin{aligned}
\sum_{\{x\}} \exp\{-\beta u\} &\geq \exp\left\{-\beta \sum_{\{x\}} u\right\} \\
\sum_{\{x\}} \exp\{-\beta E(x)\} &\geq \exp\left\{-\beta \sum_{\{x\}} E(x)\right\} \\
&\geq \exp\left\{-\beta \sum_{\{x\}} \sum_{c} V_c(x)\right\} \\
&\geq \exp\left\{-\beta U(x)\right\}
\end{aligned} \tag{4.21}$$

The prior energy with the potential function g is

$$U(f, l) = \sum_{i \in \mathcal{S}} \sum_{j \in \mathcal{N}_i} g(f_i, f_j, l_{i,j}) \tag{4.22}$$

Figure 4.5 illustrates the use of line processes in a two-dimensional lattice that typically represents the pixels in a digital image whereas Figure 4.6 illustrates the use of line processes in the three-dimensional lattice that represents the voxels in a volumetric image that is created in tomographic imaging.

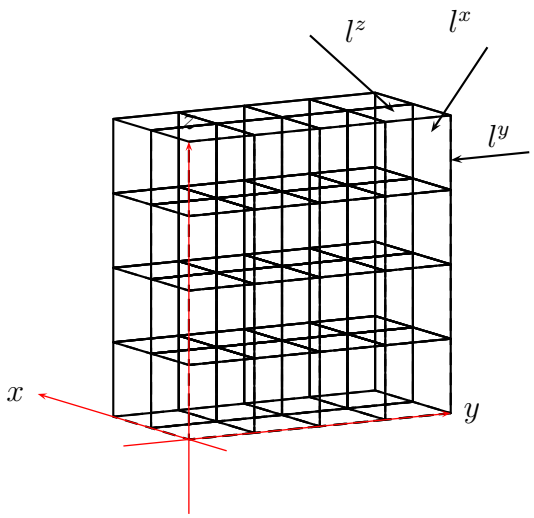


Figure 4.6: 3-D Line Processes

4.5 Roughness penalties for MRFs

Roughness penalties can be expressed as prior probabilities or an energy function $U(f)$ that measures the degree to which a smoothness assumption is violated by a configuration f. For spatially continuous MRFs as are found with 2-D and 3-D imagery, the roughness penalty involves derivatives.

A first-order difference is used to approximate a first derivative in a one-dimensional sample and summation approximate an integral. In the case of the *string*, the energy function may be expressed as

$$U(f) = \int [f(x)]^2 \, dx \tag{4.23}$$

A first-order difference is used to approximate a first derivative in a one-dimensional sample and summation approximates an integral. The discrete form of this energy takes the form

$$U(f) = \sum_i [f_i - f_{i-1}]^2 \, dx \tag{4.24}$$

Blake and Zisserman (8) introduced weak continuity constraints to model piecewise smooth functions in imagery. For the *weak string*, the energy function

is defined as

$$U(f) = \lambda^2 \int u'^2 dx + \alpha \sum_{1}^{N} l_i \tag{4.25}$$

Expressed as a sum of clique potentials, Equation 4.24 becomes

$$U(f) = \sum_{\{c\} \in \mathcal{C}} V_c(f) = \sum_{i \in S} \sum_{j \in \mathcal{N}_i} V_2(f_i, f_j) \tag{4.26}$$

where $\mathcal{C}$ consists of only pairwise cliques and

$$V_c(f) = V_2(f_i, f_j) = \frac{1}{2}(f_i - f_j)^2 \tag{4.27}$$

The two-dimensional equivalent of the string is the *membrane*, which is expressed as

$$\int \int \left\{ \left[f_x(x,y)^2 + f_y(x,y)^2 \right] \right\} \, dx \, dy \tag{4.28}$$

Numerous researchers, including Grimson (45), Blake and Zisserman, (8), Terzopoulos (92; 93) and Lee *et al* (67) have studied the use of weak membranes for the continuous restoration of surfaces or the regularization of tomographic reconstructions..

Similarly, the prior energy $U(f)$ can be designed for planar surfaces. For the *rod*, the energy function may be expressed as

$$U(f) = \int [f''(x)]^2 dx \tag{4.29}$$

In the discrete case, the second-order difference that approximates the second-order derivative may be expressed as

$$U(f) = \sum_{i} \left[f_{i+1} - 2f_i + f_{i-1} \right]^2 dx \tag{4.30}$$

The variational approach to energy minimization in plates was presented in (20, p. 250) and cited in (45, p. 235). The potential energy is expressed in terms of the principle curvatures of a three-dimensional surface, κ_1, κ_2, the mean curvature J, and

the Gaussian curvature K. The potential energy is defined by the curvatures of the plate:

$$\begin{aligned} U &= A\left(\frac{1}{\kappa_1^2}\frac{1}{\kappa_2^2} + \frac{2B}{\kappa_1\kappa_2}\right) \\ &= \left(\kappa_1^2 + \kappa_2^2\right) - 2\kappa_1\kappa_2 + 2B\kappa_1\kappa_2 \\ &= AJ^2 - 2(A-B)K. \end{aligned}$$

where A, B are constants determined by the material properties of the plate and J and K are

$$\begin{aligned} J &= \nabla^2 f = f_{xx} + f_{yy} \\ K &= f_{xx}f_{yy} - f_{xy}^2 \end{aligned} \tag{4.31}$$

Equation 4.31 may be written

$$U = \frac{1}{2}c\int\int_R [(f_{xx}^2 f_{yy}^2 - 2(1-\mu)(f_{xx}f_{yy} - f_{xy}^2)]\, dx\, dy$$

which is the potential energy of deformation of the plate due to bending.

For a quadratic surface, the third-order derivative is zero and the prior energy may be written as

$$U(f) = \int [f'''(x)]^2 dx \tag{4.32}$$

In the discrete case, the third-order difference that approximates the second-order derivative may be expressed as

$$U(f) = \sum_i \left[f_{i+2} - 3f_{i+1} + 3f_i - f_{i-2}\right]^2 dx \tag{4.33}$$

Geman and Reynolds (36) used third order differences to represent higher-order cliques in two dimensions.

The second-order smoothness priors may also be extended to two dimensions, creating the *plate*, which may be expressed as either the quadratic variation.

$$U(f) = \int\int \left\{\left[f_{xx}(x,y)\right]^2 + 2\left[f_{xy}(x,y)\right]^2 + \left[f_{yy}(x,y)\right]^2\right\}\, dx\, dy \tag{4.34}$$

or the squared Laplacian

$$U(f) = \int \int \left\{ f_{xx}(x,y)^2 + f_{yy}(x,y)^2 \right\} \, dx \, dy \tag{4.35}$$

For penalized tomographic reconstruction, surface restoration approaches may be inappropriate and the biharmonic equation may be more suitable.

$$\begin{aligned} U(f) = \int \int \big\{ & \left[f_{xx}(x,y,z) \right]^2 + \left[f_{yy}(x,y,z) \right]^2 + \left[f_{zz}(x,y,z) \right]^2 + \\ & 2 \left[f_{xy}(x,y,z) \right]^2 + 2 \left[f_{xz}(x,y,z) \right]^2 + 2 \left[f_{yz}(x,y,z) \right]^2 \, dx \, dy \, dz \end{aligned} \tag{4.36}$$

or the squared Laplacian

$$U(f) = \int \int \left\{ f_{xx}(x,y,z)^2 + f_{yy}(x,y,z)^2 + f_{zz}(x,y,z)^2 \right\} \, dx \, dy \, dz \tag{4.37}$$

Recall that Terzopoulos formulated multivariate smoothness constraints based on spline approximations defined in Equations 3.33 – 3.36 in Section 3.2. In two dimensions, the spline approximations are proportional to the small deflection bending energy of a thin plate with zero Poisson ratio.

4.6 Line process approximation

A potential function $g(\cdot)$ is often a penalty against irregularity that is meant to enforce smoothness; it corresponds to the prior clique potentials in MRFs. In Tikhonov regularization, the potential function takes a quadratic form

$$g_q(\eta) = \eta^2 \tag{4.38}$$

When a line process model is used to preserve spatial discontinuity within an image, it is typically incorporated in the potential function to simplify the energy minimization process. Blake and Zisserman (8) introduced the truncated quadratic potential

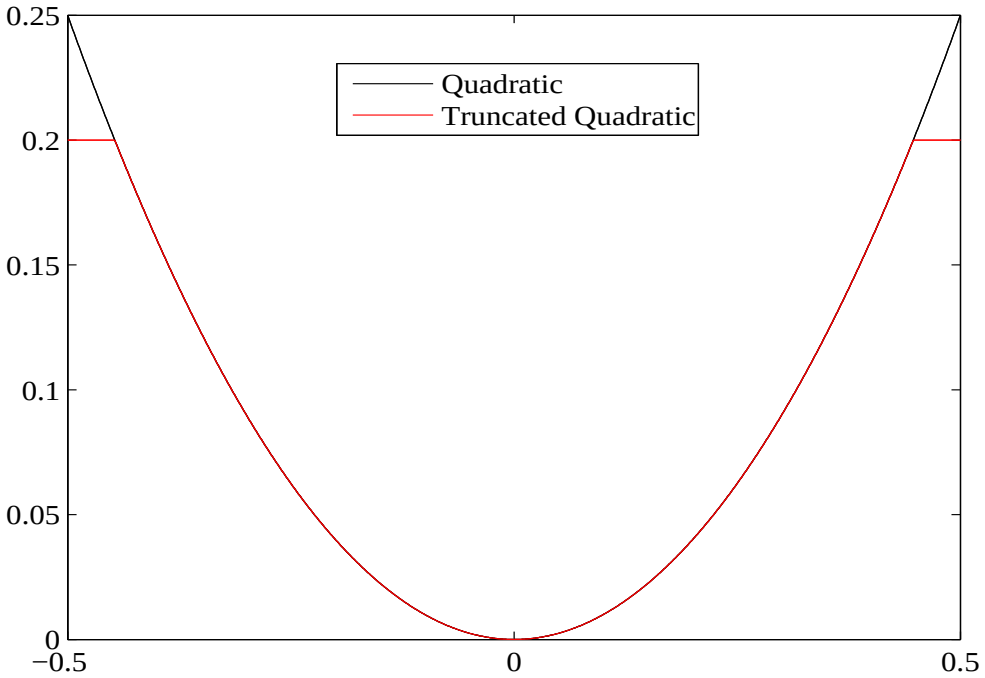


Figure 4.7: The Truncated Quadratic

function

$$g_\alpha(\eta) = \min\{\eta^2, \alpha\} \tag{4.39}$$

Shown in Figure 4.7, this potential function has a parameter p into $g_\alpha(\eta)$ to control the convexity of E, obtaining $g_\alpha^{(p)}(\eta)$. The parameter p varies from 1 to 0, which corresponds to the variation from a convex approximation of the function to its original form.

Yuille (100) performs the conversion using the Hopfield approach which models the binary line processes l_i as continuous variables $\bar{l}_i$ in the range $[0, 1]$. Each line process is related to an internal variable by a sigmoidal function

$$\bar{l}_i = \frac{1}{1 + \exp\{-\nu_i/\tau\}} \tag{4.40}$$

with τ as the parameter whereby $\lim_{\tau \to 0} \bar{l}_i = l_i$. The energy with this treatment is the Gibbs distribution

$$P(f \mid l) = \frac{1}{Z} \exp^{-\frac{1}{T}U(f,l)} . \tag{4.41}$$

Yuille (100) showed that at stationary points where $dE/dt = 0$, $v_i = \lambda[f_i - f_{i-1}]^2 - \alpha)$ and hence the approximated line process variables

$$\bar{l}_i = \frac{1}{1 + \exp(-\nu/\tau)} \tag{4.42}$$

This gives the effective potential function as

$$g_{\alpha,\tau}(\eta) = -\frac{\tau}{2} \ln[1 + \exp\left[-\lambda([f_i - f_{i-1}]^2 - \alpha)/\tau\right] \tag{4.43}$$

As the temperature decreases toward zero, $\bar{l}_i$ approaches $l_i : \lim_{\tau \to 0^+} \bar{l}_i = l_i$.

Geiger and Girosi (35) approximated the line process using mean field theory. They introduced a parameter β into the Gibbs distribution, giving an approximate posterior probability

$$P_\beta(f, l \mid d) = \frac{1}{Z_\beta} \left[P(f, l \mid d)\right]^\beta \tag{4.44}$$

Using the saddle point approximation method, they derive mean field equations which yield the approximated line process variables which are identical to Equation 4.43. The solution is in the limit $f^* = \arg\max_{f,l} \lim_{\beta \to \infty} P_\beta(f, l \mid d)$.

Using the *minimum description length* principle, Leclerc (66) designed an energy function for restoration of a piecewise constant image f from noisy data d.

$$U(f \mid d) = \sum_{i=1}^{M} [f_i - d_i]^2 + \lambda \sum_{i=2}^{M} [1 - \delta(f_i - f_{i-1})] \tag{4.45}$$

where $\delta(\cdot) \in \{0, 1\}$ is the *Kronecker delta function* and λ is a constant weighting factor.

These line process approximations often introduce non-convex energy functions which complicate energy minimization. The problem can be solved by searching over all configurations of (f, l) but this is impractical due to the large number of configurations. With the truncated quadratic, Blake and Zisserman (8, p. 114) have shown that the solution for the line process l_{ij} reduces to

$$l_{i,j}^{v} = \begin{cases} 0, & V_{i,j}^{h} \leq \frac{\alpha}{\lambda} \\ 1, & V_{i,j}^{h} > \frac{\alpha}{\lambda} \end{cases} \tag{4.46}$$

$$l_{i,j}^{h} = \begin{cases} 0, & V_{i,j}^{v} \leq \frac{\alpha}{\lambda} \\ 1, & V_{i,j}^{v} > \frac{\alpha}{\lambda} \end{cases} \tag{4.47}$$

for the weak membrane (39; 40; 67) where $l_{i,j}^{h}$ and $l_{i,j}^{v}$ represent the horizontal and vertical line processes, respectively. A pictorial representation of these 2-D line processes is given in Figure 4.5. The line processes $l_{i,j}^{h}$ and $l_{i,j}^{v}$ may also be denoted as $l_{i,j}^{x}$ and $l_{i,j}^{y}$ to denote explicitly edge transitions in the x and y-axes, respectively. For the weak plate and thin-plate spline, a single line process is associated with each voxel:

$$l_{i,j} = \begin{cases} 0, & V_{i,j} \leq \frac{\alpha}{\lambda} \\ 1, & V_{i,j} > \frac{\alpha}{\lambda} \end{cases} \tag{4.48}$$

and

$$l_{i,j,k} = \begin{cases} 0, & V_{i,j,k} \leq \frac{\alpha}{\lambda} \\ 1, & V_{i,j,k} > \frac{\alpha}{\lambda} \end{cases} \tag{4.49}$$

For the weak membrane, the energy function $V_{i,j}$ is calculated as

$$V_{i,j} = f_x^2(i,j) + f_y^2(i,j). \tag{4.50}$$

For the weak plate, it is calculated as

$$V_{i,j} = f_{xx}^2(i,j) + 2\left[f_{xy}^2(i,j)\right] + f_{yy}^2(i,j). \tag{4.51}$$

For the thin-plate spline, $V_{i,j,k}$ is calculated as

$$\begin{aligned} V_{i,j,k} = f_{xx}^2(i,j,k) + f_{yy}^2(i,j,k) + f_{zz}^2(i,j,k) + \\ 2\left[f_{xy}^2(i,j,k)\right] + 2\left[f_{xz}^2(i,j,k)\right] + 2\left[f_{yz}^2(i,j,k)\right]. \end{aligned} \tag{4.52}$$

Lee *et al* (70) introduced the notation $V_{i,j}$ instead of $U_{i,j}$ as a convenient method of representing both the weak membrane and higher-order clique potentials in a unified manner. Using this notation, the new prior energy becomes

$$U(f) = \sum_{i,j,k} \phi^*\left(\sqrt{V_{i,j,k}}\right) \tag{4.53}$$

where

$$\phi^*\left(\sqrt{V_{i,j,k}}\right) = \begin{cases} \lambda V_{i,j,k}, & \lambda V_{i,j,k} \leq \alpha \\ \alpha, & V_{i,j,k} > \alpha. \end{cases} \tag{4.54}$$

Figure 4.7 shows $\phi^*\left(\sqrt{V_{i,j,k}}\right)$ as a truncated quadratic function. The line process model assumes piecewise smoothness whereby the smoothness constraint is switched off at points where the magnitude of the signal derivative exceeds a certain threshold. It is defined on a lattice rather than on a continuous domain. A continuous interval is quantized into m uniformly-spaced points $x_1, x_2, \ldots, x_m$ so that $f_i = f(x_i), d_i = d(x_i)$ and $\chi_i = \chi(x_i)$. A set of binary variables $l_i \in \{0,1\}$ is introduced to represent the line processes into the smoothness term. If $w_i = w_i(x_i)$ takes the neighboring points $i-1$ and i takes on a value in $\{0,1\}$ then $i-1$ and i are related by $l_i = 1 - w_i$ then $i-1$ and i are related by $l_i = 1 - w_i$ The state $l_i = 1$ of the line process variable indicates that a discontinuity is detected between the neighboring point $i-1$ and i; conversely, the state $l_i = 0$ indicates that the value between the two points is continuous. Each

activation of a line process variable is penalized by a quantity $\beta\alpha$ which gives the regularizer

$$U(f,l) = \beta \sum_{i=2}^{m} [f_i - f_{i-1}]^2 [l - l_i] + \beta\alpha \sum_{i=2}^{m} l_i. \tag{4.55}$$

The energy for the line process model is

$$E(f,l) = U(f,l \mid d) = \sum_{i=1}^{m} \chi_i [f_i - d_i]^2 [l - l_i] + \alpha \sum_{i=2}^{m} l_i. \tag{4.56}$$

Finding $f^* \in \mathbb{R}^2$ and $l^* \in 0,1$ such that $U(f,l \mid d)$ is minimized is a mixture of real and combinatorial optimization. Algorithms for this can be classified as either stochastic (37) or deterministic (8; 35; 100). Some annealing techniques are often combined to obtain global solutions. In stochastic approaches, f and l are updated according to some probability distribution parameterized by a temperature parameter. For example, Geman and Geman (37) used simulated annealing with the Gibbs sampler to find the global MAP solution. Deterministic approaches often use some classical gradient-based methods. Before these can be applied, the combinatorial minimization problem has to be converted into one of real minimization. By eliminating the line process, Blake and Zisserman (8) converted the previous minimization problem into one which minimized the following function containing only real variables

$$U(f \mid d) = \sum_{i=1}^{m} \chi_i [f_i - d_i]^2 + \alpha \sum_{i=2}^{m} g_\alpha (f_i - f_{i-1}). \tag{4.57}$$

where $g_\alpha(\cdot)$ is the truncated quadratic function given in Equation 4.39. They introduced a parameter p into $g_\alpha(\eta)$ to control the convexity of E, obtaining $g_\alpha^{(p)}(\eta)$. The parameter p varies from one to zero, which corresponds to the transformation from a convex approximation of a function to its original form. Black and Rangarajan (7) describe the properties of a number of commonly-used potential functions, some of which are shown in Table 1.1.

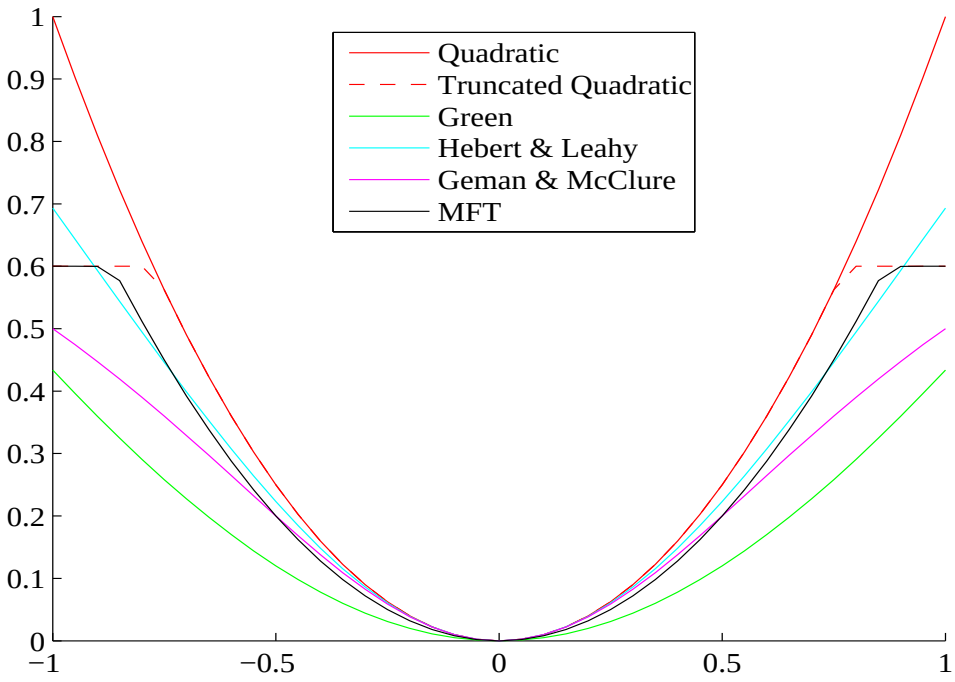


Figure 4.8: Potential functions

Chapter 5

Image Reconstruction Techniques

Image reconstruction techniques based on algebraic or statistical approaches allow the physics of photon emission and detection to be modeled accurately and can incorporate anatomical or geometric information to improve image quality. Given the size of the system matrix that models the scanner geometry and other Factors such as detector efficiency, iterative methods are typically used to solve the reconstruction problem. An initial estimate of radiotracer uptake is made, and that estimate is refined until some stopping criteria are met. Typically, computation will proceed until the error norm of the estimated solution has reached some predefined tolerance.

Independently, Shepp and Vardi (89) and Lange and Carson (63) proposed a maximum likelihood likelihood algorithm for expectation minimization (MLEM) to reconstruct tomographic images.

(1) E-step: Form $Q(f \mid f^{(k)}) = E\left[\log P(f, d \mid f^{(k)}) \mid d, f^{(k)} - \beta R(f^{(k)})\right]$

(2) M-step: Find $f^{(k+1)} = \arg\max_x Q(f \mid f^{(k)})$.

In the case of Poisson data, these steps combine into a single step, which can be expressed as a multiplicative update rule:

$$f_i^{(k+1)} = f_i^k \sum_d \frac{y_d^*}{f_d^*} p_{bd} \tag{5.1}$$

where $f_d^* = \sum_i f_i^k p_{bd}$ and $\sum_d p_{bd} = 1$. MLEM has three important properties, nonnegativity, monotonicity and count preservation, but its convergence can be slow. As reconstruction is an ill-posed problem, the reconstruction can take on an unsmooth quality as the solution is approached. A simple technique for resolving this problem is to initialize the image with a smooth estimate and terminate the algorithm prior to convergence. As described in Chapter 4, more sophisticated methods of dealing with *regularization* include the use of penalty terms based on Markov random fields. These penalty terms change the form of expectation maximization as it is likely that the M-step no longer has a closed form solution. In this case, the *Generalized Expectation Maximization* (GEM) algorithm provides an approximate solution by ensuring that the update does not decrease the likelihood of the estimate. In this case, GEM takes the form

(1) E-step: Form $Q(f \mid f^{(k)}) = E\left[\log P(f, d \mid f^{(k)}) \mid d, f^{(k)} - \beta R(f^{(k)})\right]$

(2) M-step: Find $Q(f^{(k+1)} \mid f^{(k)}) \geq Q(f \mid f^{(k)})$.

For a prior with a uniform distribution, MAP-GEM reduces to the MLEM algorithm. The penalized maximum likelihood equation for GEM (PML-GEM) may be expressed as :

$$f^{k+1} = \arg\max \log P(d \mid f^k) - \frac{\beta}{T} \sum_{ijk} \sum_{l \in L} V_c(f_{ijk}^k) \tag{5.2}$$

$$f_{ijk}^{k+1} : \frac{\partial}{\partial f_{ijk}} \log P(d \mid f^k) - \frac{\beta}{T} \sum_{ijk} V_c(f_{ijk}^k) = 0 \tag{5.3}$$

Hebert and Leahy (48) and Mumcuoğlu *et al* (82) have used conjugate gradient techniques in conjunction with MAP-GEM reconstructions. The former used the *Polak-Ribière* conjugate gradient algorithm and coordinate ascent with a backtracking line search to determine the optimal step size whereas the latter used a Newton-Raphson procedure to determine the step size. They compute maximum *a posteriori* (MAP) estimates or a penalized maximum likelihood estimate as defined in Equation 5.2 using different 3-D Gibbs priors, including a quadratic prior and the Hebert-Leahy prior shown in Table 1.1.

There have been several attempts to accelerate the convergence of image reconstruction algorithms. The ordered subsets EM (OSEM) algorithm of Hudson and Larkin (51) is a modification of the MLEM algorithm in which each voxel update uses only a subset of the observed data. Typically, each subset will consist of a group of projections. The subsets should be chosen so that an emission from each voxel has equal probability of being detected in each of the subsets. The grouping of projections within the subsets will affect both the convergence rate and the sequence of generated images. To avoid directional artifacts, the projections are usually chosen to have maximum separation in angles. Due to more rapid convergence, this approach has been used extensively in clinical settings. One disadvantage of this algorithm is that it limits the types of priors or penalty terms that can be incorporated into the reconstruction.

Some algorithms transform the maximum likelihood problem using *surrogate* functions such as paraboloidal surrogates algorithms of Erdoğan and Fessler (27). Based on the principle of *optimization transfer*, the algorithm allows the optimization of one function to be approximated by the maximization of a simpler problem. The optimum curvature of an approximating parabola is calculated. The penalized log-likelihood function for the Poisson process is approximated by two quadratic functions, one for the log-likelihood term and another for a regularizing penalty term. The *majorization principle* guarantees convergence by ensuring that the quadratic

function modeling the penalty term is always less than or equal to the original penalty function. A coordinate descent algorithm is then used to find the optimal solution.

Kaufmann (58) and others have restated the update rule of the EM algorithm rule as

$$\begin{aligned} f_i^{(k+1)} &= f_i^k + f_i^k \frac{\partial l(f_i^k)}{\partial f_i^k} \\ &= f_i^k (1 + \nabla l(f_i^k)). \end{aligned}$$

so that it may be considered as a scaled steepest descent algorithm with the distance to the nonnegativity constraint used as a scaling factor. Convergence can be accelerated by using line searches and other optimization techniques (75). Line searches have a long tradition in optimization and based on searching a new iterative point along a descent direction at each iteration. Generally, line search methods take the form

$$f_{k+1} = f_k + \alpha_k d_k, k = 0, 1, 2, \ldots, \tag{5.4}$$

where d_k is a descent direction of $f(x)$ at f_k and α_k is a step size.

Another approach by Kaufmann (59), Fessler (28) and others is to treat the reconstruction problem as a sparse, overdetermined system of equations and employ traditional minimization techniques to find $\phi(\hat{f})$ where

$$\hat{f} = \arg\min_f \|d - Af\|_2^2 + \beta R(f). \tag{5.5}$$

Using these techniques requires that nonnegativity be handled separately. This is often accomplished with restricted or bent line searches (59), penalty functions (82), interior-point methods (55) or active sets (59). Gradient-based techniques can then be used to solve these penalized systems, but some expertise must be used to select the appropriate preconditioner and penalty measures. Mumcuoğlu *et al* (81) applied active set techniques to 3-D image reconstruction, but this approach has limited

appeal because it requires reinitialization when the set of active variables changes. Moré and Toraldo (79; 80) and Bierlaire, Toint and Tuyttens (6) have applied an active set approach using a gradient projection to combine unconstrained optimization over a restricted set of variables.

Several preconditioning strategies have been proposed to solve image reconstruction problems using the large, sparse matrices that comprise the system model. problems. Clinthorne *et al* (19) consider preconditioning as a filtering operation in the frequency domain and construct a high-pass filter from a point-spread function of $A^T A$, which could be dense, by taking a single pixel of the center of the field of view. Chinn and Huang (17) derive a class of preconditioners based on the discrete Radon transform. Lastly, Fessler and Booth (29) demonstrate that the circulant preconditioners used typically in imaging problems are inappropriate for nonuniform noise variance as is often found in Poisson statistics for emission tomography, and they propose a a preconditioner that better approximates the Hessian matrix for shift-variant imaging problems.

If the Hessian $\nabla^2 f(f_k)$ is available and invertible, then $d_k = \nabla^2 f(f_k)$ leads to the Newton method while $-d_k$ results in the steepest descent method. The search direction d_k is generally required to satisfy

$$d_k^T \nabla f_k < 0 \tag{5.6}$$

which guarantees that d_k is a descent direction of $f(x)$ at x_k. In order to guarantee the global convergence, d_k may be required to satisfy the sufficient decrease condition

$$f(x_k + \alpha d_k) \leq f(x_k) + c_1 \nabla f_k^t d_k \tag{5.7}$$

where $c > 0$ is a constant. Instead of Equation 5.7, the angle property is often used in proving the global convergence of related line search methods, namely the angle θ_k

between d_k and the steepest descent direction $-\nabla f_k$,

$$\cos\theta_k = \frac{-\nabla f_k^T d_k}{||\nabla f_k||\,||d_k||} \tag{5.8}$$

Once the descent direction d_k is obtained, an appropriate step size is determined along the descent direction and the iteration continues. There are many approaches to find an available step size. It is well known that the exact line search is time-consuming, therefore inexact line search rules are used frequently. Examples include the Armijo, Goldstein and Wolfe rules where the Armijo rule is the easiest to implement for practical computation (84). Convergence analysis on line search methods can be found in the literature (75; 84). Another approach uses the *conjugate gradient* method, which has been extended to nonconvex functions such as some of the penalty functions shown in Table 1.1 (33; 42). If f is a strongly convex quadratic function and α_k is the exact minimizer, this algorithm reduces to the linear conjugate gradient algorithm.

5.1 Graduated Non-Convexity Algorithms (GNC)

Blake and Zisserman (8) derived the Graduated Non-Convexity algorithm (GNC) to minimize the cost function arising from the elimination of binary line processes from weak mechanical priors (e.g., weak membrane or weak plate). They devised a series of approximations designed to model a non-convex penalty function. They applied a gradient descent algorithm to the approximations. Geiger and Girosi (35) used mean field theory to approximate the line processes and demonstrated that GNC was a form of deterministic annealing. Bedini, Gerace and Tonazzini (3) proposed a method that permits self-interactions among line processes by representing them as a sigmodial function of voxel gradients similar to that of (67; 70; 40).

Blake and Zisserman (8, page 119) use the following energy functions for the weak plate:

$$\begin{aligned} F &= D + \sum_{ij} g_{\beta,\mu^2}(f_{i,j-1} + f_{i,j+1} + f_{i-1,j} + f_{i+1,j} - 4f_{i,j}) \\ F &= D + \sum_{ij} g_{\beta,\mu^2}\left(\sqrt{V_{ij}}\right) \end{aligned} \tag{5.9}$$

Using a Laplacian kernel, these functions may be extended to three dimensions:

$$\begin{aligned} F &= D + \sum_{ijk} g_{\beta,\mu^2}(f_{i,j-1,k} + f_{i,j+1,k} + f_{i-1,j,k} + f_{i+1,j,k} - f_{i,j,k-1} - f_{i,j,k+1} - 6f_{i,j,k}) \\ F &= D + \sum_{ijk} g_{\beta,\mu^2}\left(\sqrt{V_{ijk}}\right) \end{aligned} \tag{5.10}$$

Blake and Zisserman (8) constructed a set of functions $\bar{F}(f)$ which approximate $F(f)$.

$$g^{(p)}(t) = \begin{cases} \lambda^2 t^2, \text{if } |t| < q \\ \alpha - c(|t| - r)^2/2, \text{if } q \leq |t| < r \\ \alpha, \text{if } |t| \geq r \end{cases} \tag{5.11}$$

where

$$c = \frac{c^*}{p}, r^2 = \alpha\left(\frac{2}{c} + \frac{1}{\lambda^2}\right), q = \frac{\alpha}{\lambda^2 r} \tag{5.12}$$

and showed that, when $0 < c^* < \frac{1}{4}$, $F^{(1)}$ is convex. The selection of the parameter c^* is based upon the calculation of the largest eigenvalue ν_{max} of the matrix $Q^T Q$ where Q is a matrix that describes a linear transformation that is a discrete approximation of the differential operator that defines the energy of the fitted surface. When Q is a circulant matrix it can be represented by convolution. The eigenvalues of a circulant matrix are explicitly known (21, page 73)

$$\nu_{mno} = \sum_{rst} C_{rst}^2 \exp(2\pi imr/N) \exp(2\pi ins/N) \exp(2\pi iot/N), 0 \leq m, n, o \leq N. \tag{5.13}$$

The membrane energy differs from other energies in that its energy leads to a sum of two terms in the Hessian, expressed in terms of two circulants Q, R with corresponding 2D convolutions C, D. Since $Q^TQ = Q^2$, the eigenvalues of Q^TQ are

$$\nu_{mn} = 4 - 2\cos(2\pi m/N) - 2\cos(2\pi n/N) \tag{5.14}$$

so that $\nu_{max} = 8$.

The square Laplacian energy for a plate yields a convolution looks like this:

$$C = \begin{pmatrix} \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & 0 & -1 & 0 & \cdots \\ \cdots & -1 & 4 & -1 & \cdots \\ \cdots & 0 & -1 & 0 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix} \tag{5.15}$$

As before, $Q^TQ = Q^2$ so the eigenvalues of Q^TQ are squared to produce

$$\nu_{mn} = (4 - 2\cos(2\pi m/N) - 2\cos(2\pi n/N))^2 \tag{5.16}$$

so that $\nu_{max} = 8^2 = 64$.

For the minimization of $F^{(p)}$, the nth iteration is

$$u_l^{(n+1)} = u_l^n - \omega \frac{1}{T_l} \frac{\partial F^{(p)}}{\partial u_l} \tag{5.17}$$

where $0 < \omega < 2$ is the relaxation parameter, which governs the speed of convergence, and T_l is an upper bound on the second derivative

$$T_l \geq \frac{\partial^2 F^{(p)}}{\partial u_l^2} \forall u \tag{5.18}$$

An alternative is to consider the 3D discrete Laplacian on a regular grid with a homogeneous Dirichlet boundary condition. Using the Kronecker sum:

$$L = D_{xx} \oplus D_{yy} \oplus D_{zz} = D_{xx} \otimes I \otimes I + I \otimes D_{yy} \otimes I + I \otimes I \otimes D_{zz}, \tag{5.19}$$

where D_{xx}, D_{yy} and D_{zz} are 1D discrete Laplacians in each of the three axes, and I are identities are appropriate sizes. The resulting eigenvalues are

$$\lambda_j = -\frac{4}{h^2} \sin\left(\frac{\pi j}{2(n+1)}\right)^2, 1 \leq j \leq 3. \tag{5.20}$$

The boundary points lie just outside of the grid where the homogeneous Dirichlet boundary condition is imposed.

The resulting algorithm is based on a form of nonlinear successive over-relaxation (SOR) where the update equation is given by (8, page 157):

$$f_{ijk}^{(n+1)} = f_{ijk}^{(n)} - \omega \left\{ 2 \left(f_{ijk}^{(n)} - d_{ijk} \right) \sum_{ijk} g_k^*(f) \right\} \tag{5.21}$$

5.2 Hyperparameter Selection

A major challenge with MAP-GEM is the selection of one or more hyperparameters which control the smoothing of the prior. Typically, these hyperparameters are initialized at the beginning of the reconstruction process. Depending upon the hyperparameter, its value may remain static throughout the reconstruction or it may be modified based on interim results during the reconstruction. The hyperparameters control the trade-off between noise reduction (i.e., variance) and resolution (i.e., bias) in reconstructed images. When the hyperparameter is too large, the smoothing effects adversely affect image resolution, but when it is too small, the reconstruction mirrors MLEM, which may produce images which have high contrast and high noise levels.

There are two basic approaches to hyperparameter selection. The first treats the hyperparameter as a regularization parameter and uses generalized cross-validation (GCV) (96), L-curves (46; 47; 101) or χ^2 goodness of fit methods (94; 49). The second treats selection as a separate estimation problem and uses MLEM or some other estimation technique to choose a hyperparameter.

The main idea behind generalized cross-validation is to omit a portion of the data set, estimate the model for various values of the hyperparameter and select the value that best predicts the omitted data. Its disadvantages include a difficulty in obtaining a numerical solution and its lack of accuracy in the presence of correlated noise. In high-dimensional problems such as computed tomography, the method is seen as impractical due to its high computational requirements.

Hansen and O'Leary (47) proposed the use of an *L-curve* to select hyperparameters. The L-curve is a plot of the Gibbs energy versus the log-likelihood for the MAP estimate at each value of the hyperparameter.

5.3 Line Process Formation

The transfer function is a semilinear sigmoid, or logistic, function which takes the form

$$f(x) = \frac{1}{1 + \exp(-bx)}. \tag{5.22}$$

This function is an increasing function which is concave if $bx > 0$ and convex if $bx < 0$. This sigmoid function is used to raise a specific intensity range. The intensity range is described by its center β and its width α. The target intensity range is given by the interval [min, max].

$$f(x) = \frac{max - min}{1 + \exp(-bx)}. \tag{5.23}$$

Typically, the transfer function is a semilinear sigmoid function, as shown in equation 5.22. In a narrow range around 0, the function is almost linear. Seen more

Table 5.1: A selection of sigmoid transfer functions

	Function
Logistic sigmoid	$\Phi(u) = \frac{1}{1+\exp(-u)}$
Bipolar	$\Phi(u) = \frac{2}{1+\exp(-u)} - 1$
Hyperbolic Tangent sigmoid	$\Phi(u) = \tanh(u) = \frac{2}{1+\exp(-2u)} - 1$

globally, the function acts as a Heaviside step-function, which is often called a soft threshold function. A small selection of transfer functions is presented in Table 5.1.

Researchers have suggested that the line processes should not be formed until a certain number of iterations of MLEM have concluded. No guidelines have been established, however, for determining an appropriate number of iterations.

As the formation of line processes is greatly affected by hyperparameter values, upper and lower bounds must be established. The lower bound for α is calculated as follows:

$$\begin{aligned} e^{-\beta\gamma(u-\alpha)} &< \tau \\ -\beta\gamma(u-\alpha) &< \log\tau \\ -\beta\gamma u + \beta\gamma\alpha + &< \log\tau \\ \beta\gamma\alpha &< \log\tau + \beta\gamma u \\ \alpha &< \frac{\log\tau + \beta\gamma u}{\beta\gamma} \end{aligned} \tag{5.24}$$

Similarly, the upper bound for α is calculated as follows:

$$\begin{aligned} e^{-\beta\gamma(u-\alpha)} &> 1-\tau \\ -\beta\gamma(u-\alpha) &> \log(1-\tau) \\ -\beta\gamma u + \beta\gamma\alpha + &> \log(1-\tau) \\ \beta\gamma\alpha &> \log(1-\tau) + \beta\gamma u \\ \alpha &> \frac{\log(1-\tau) + \beta\gamma u}{\beta\gamma} \end{aligned} \tag{5.25}$$

In addition, the temperature T has an lower bound to prevent the variable being optimized from attaining a negative value:

$$T >> \max\left(1, \frac{1}{\alpha}\right). \tag{5.26}$$

5.4 Annealing Techniques for Energy Minimization

In condensed matter physics, annealing denotes a physical process in which a solid in a heat bath is heated by increasing the temperature of the bath to a maximum value. The particles of the solid randomly arrange themselves in the liquid phase and subsequent cooling allows individual particles to arrange themselves in a low energy ground state.

At each temperature T, the solid reaches thermal equilibrium, which is characterized by a probability of being in a state with energy E given by the *Boltzmann distribution*, which is also known as the Gibbs distribution.

$$P(E) = \frac{1}{Z(T)} \exp \frac{-E}{k_B T}. \tag{5.27}$$

As the temperature decreases, this distribution concentrates on the states with the lowest energies. As the temperature approaches zero, only those states with minimum energy have a non-zero probability of occurrence. If the temperature reduction (i.e., cooling) is too rapid, the process can arrive at local minima, which corresponds to solid or metastable structures in the physical analogy. Gindi *et al* (40) present results on process of *quenching* an annealing process, which is analogous to lowering the temperature instantaneously.

Simulated annealing (95), also known as stochastic relaxation (37), does not depend upon the initial estimate to arrive at a solution as do some algebraic techniques, but it may take a long time to obtain thermal equilibrium at low

temperatures. Although guaranteed to converge to a global minimum, the amount of time required to find a solution makes the approach impractical for most problems. Deterministic annealing is a *continuation method* as described in (85, pages 230–234), which approximates a function through a sequence of smooth functions is indexed by T. The Gibbs distribution may be transformed as follows:

$$P_T(d \mid f) = \frac{1}{Z(f;T)} \exp(-TU(f)), \tag{5.28}$$

where $Z(f;T)$ is the partition function parameterized by T. This parameterization is described in Section 5.5 following the examples of (67; 39; 40; 70).

5.5 A Bayesian approach

Using Bayes Theorem, the *a posteriori* conditional probability distribution of source intensities F given the projection data D may be written as

$$P(f, \mid d) = \frac{P(d \mid f)P(f)}{P(d)} \tag{5.29}$$

where $P(d \mid f)P(f)$ is the likelihood function and $P(f)$ is the *a priori* probability distribution of F.

With the introduction of lines processes L, Equation 5.29 can be expressed as

$$P(f, l \mid d) = \frac{P(d \mid f, l)P(f, l)}{P(d)} \tag{5.30}$$

Given the posterior distribution in Equation 5.30, a possible estimator is the Maximum *a posteriori* (MAP) estimate. Since $P(d)$ is a constant for fixed d, $P(f, l \mid d)$ is proportional to the joint distribution

$$P(f \mid d) \propto P(f, d) = P(d \mid f)P(f) \tag{5.31}$$

Then the MAP estimate is found by

$$(f^*, l^*) = \arg\max_{\{f,l\}} P(d \mid f, l)P(f, l) \tag{5.32}$$

as shown below:

$$\begin{aligned}
(\hat{f}, \hat{l}) &= \arg\max_{\{f,l\}} P(f, l \mid d) \\
&= \arg\max_{\{f,l\}} P(D \mid f, l)P(f, l) \\
&= \arg\max_{\{f,l\}} \log P(d \mid f) + \log P(f \mid l) + \log P(l) - \log P(d) \\
&= \arg\max_{\{f,l\}} \log P(d \mid f) + \log P(f \mid l) + \log P(l). \\
&= \arg\min_{\{f,l\}} -\log P(d \mid f, l) - \log P(f).
\end{aligned} \tag{5.33}$$

The term $\log P(d)$ can be eliminated since the arg max is carried out only with respect to f and l. Since the projection data D is obtained from forward projection of the source intensities F, it does not involve the line processes L. Therefore, the likelihood term $P(d \mid f, l)$ may be rewritten as $P(d \mid f)$. The prior term $P(f, l)$ is rewritten as

$$P(f, l) = P(f \mid l)P(l), \tag{5.34}$$

where the conditional probability $P(f, l)$ is interpreted as a prior on f conditioned on l and the probability $P(f, l)$ as a prior solely on l. The divisor $P(d)$ can be eliminated since the maximum is carried out with respect to only f and l.

The likelihood and prior energy may then be modeled as Gibbs distributions; the prior energy is analogous to the roughness penalty in penalized maximum likelihood as both terms represent energies based on the degree of smoothness in the reconstruction.

The prior of the joint Gibbs distribution of f and l may be written as

$$P(f, l) = \frac{1}{Z} \exp^{-\frac{1}{T}U(f,l)}. \tag{5.35}$$

The marginal posterior distribution can be written as

$$\begin{aligned} P(f \mid d) &= \sum_{l} P(f, l \mid d) \\ &= \sum_{l} \frac{P(d \mid f) P(f|l)}{P(d)} \\ &= \sum_{l} \frac{P(d \mid f)}{P(d)} \left[\sum_{l} P(f, l) \right] \\ &= \sum_{l} \frac{d \mid f)}{P(d)} P(f). \end{aligned} \tag{5.36}$$

where $P(f)$ is the marginal intensity prior, which can be evaluated using the prior distribution in Equation 5.33 and integrating out the line processes.

$$\begin{aligned} P(f) &= \sum_{\{l\}} P(f, l) \\ &= \sum_{\{l\}} \frac{1}{Z} \exp -T \left[\sum_{i,j,k} U(f_{ijk})(1 - l_{ijk}) + \alpha \sum_{ijk} l \right] \\ &= \sum_{l_{ijk}} \frac{1}{Z} \exp \left[-\beta \sum_{i,j,k} U(f_{ijk})(1 - l_{ijk}) + \sum_{ijk} \beta l_{ijk} \right] \\ &= \prod_{ijk} \left[\sum_{l=\{0,1\}} \exp \left[-T \sum_{ijk} U(f_{ijk})(1 - l_{ijk}) + \sum_{ijk} \beta l_{ijk} \right] \right] \\ &= \prod_{ijk} \left[\exp \left[-T \sum_{ijk} U(f_{ijk})(1 - l_{ijk}) + \sum_{ijk} \beta l_{ijk} \right] \right] \\ &= \exp \left[\sum_{i,j,k} \log \left[-\beta U(f_{ijk}) \right] + \exp \left[-\beta \alpha \right] \right] \\ &= \exp \left[-T \sum_{i,j,k} \left(\frac{-1}{T} \right) \log \left[-T U(f_{ijk}) \right] + \exp \left[-T \alpha \right] \right] \end{aligned} \tag{5.37}$$

As noted by Lee (68, p. 104), this derivation depends upon the identity:

$$\sum_{l \in \{0,1\}} \exp(-TAl) = 1 + \exp(-TA) = \exp \left[-T \left\{ -\frac{1}{T} \log(1 + \exp(-TA)) \right\} \right]. \tag{5.38}$$

Therefore, the marginal posterior for the deterministic annealing scheme is given by

$$\begin{aligned} P_T(f \mid d) &= \frac{1}{Z(d;T)}[-T\{E_L(f) + \sum_{i,j,k} -\frac{1}{T}\log\{\exp(-T\beta U(f_{ijk}) + \exp(-T\alpha))\}\}] \\ &= \frac{1}{Z(d;T)}[-T\{E_L(f) + \sum_{i,j,k} \phi_T(\sqrt{U_{i,j,k}})\}], \end{aligned} \tag{5.39}$$

where $\phi_T(\sqrt{U_{i,j,k}})$ is defined as

$$\phi_T(\sqrt{U_{i,j,k}}) = -\frac{1}{T}\log\left[(\{\exp(-T\beta U_{i,j,k}) + \exp(-T\alpha).\right. \tag{5.40}$$

The term $\sum_{l_{i,j,k}}$ denotes the summation over all configurations of l, where l is either 0 or 1 for all (i,j,k).

Following (68; 70), the energies for the log likelihood and the prior may be denoted as E_L and E_P, respectively. The log likelihood energy E_L may be represented as

$$E_L(f) = -\log P(d|f) = \sum_{t,\theta}\left[-d_{t,\theta}\log(\bar{d}_{t,\theta}) + \log(d_{t,\theta}!)\right] + \sum_{t,\theta}\bar{d}_{t,\theta} \tag{5.41}$$

where $\bar{d}_{t,\theta}$=$\sum_{i,j,k} A_{t,\theta;i,j,k} f_{i,j,k}$. The prior energy E_P was defined previously as $U(f)$ in Equation 4.6 and incorporated line processes using the truncated quadratic potential function.

5.6 Introduction of a Control Parameter

Lee *et al* (67) and Lee, Rangarajan and Gindi (70) use a smooth, differentiable approximation of the truncated quadratic penalty function of Blake and Zisserman (8) which is obtained using by introducing a temperature control parameter T into the potential function ϕ using continuation method where the following sequence of

smooth potential functions is indexed by T:

$$\phi_T(f) = -\frac{1}{T}\log\left[\exp\left\{-T\beta U(f)\right\} + \exp\left\{-T\beta\alpha\right\}\right]. \tag{5.42}$$

For $T = 1$, $\phi_T(f) = \phi(f)$. At large values of T, which corresponds to low temperatures in an annealing scheme, the first term in the expression dominates.

The goal is to find the minimum of the objective function by performing a minimization procedure at each temperature, which is then used as an initial value for the next temperature setting.

The truncated quadratic ϕ^* is obtained by eliminating the line process from the original energy function. The marginal prior distribution of the intensities is obtained by summing over all configurations of the line processes in the prior distribution. The resulting marginal prior is a smooth Gibbs prior and its potential function is a smooth approximation of the truncated quadratic function. The smooth potential function suggests the use of a continuation method which uses a sequence of smooth potential functions, denoted ϕ_T, indexed by the temperature parameter T. The function ϕ_T is a smooth approximation of the truncated quadratic at high temperatures, i.e., which are at low values of T, and approach the truncated quadratic at low temperatures, which are at high values of T. Despite the absence of explicit line processes in the final object function to be minimized, they are contained in the sequence of surrogate functions which approximate the truncated quadratic at limiting values of T.

5.7 Derivation of a Deterministic Annealing Algorithm

Once the MAP estimation problem has been reduced to a combination of minimizing the objective function $E_T(f)$ at each epoch T, a deterministic annealing algorithm may be devised using techniques developed by (39; 40; 70), among others. Penalized maximum likelihood (PML) can be written as

(1) E-step: Form $Q(f \mid f^{(k)}) = E\left[\log P(f, d \mid f^{(k)}) \mid d, f^{(k)} - \beta R(f^{(k)})\right]$

(2) M-step: Find $Q(f^{(k+1)} \mid f^{(k)}) \geq Q(f \mid f^{(k)})$.

The roughness penalty $R(f)$ has been added to the expectation step and GEM provides an approximate solution through an update process that ensure that the likelihood does not decrease.

The energy function has been transformed by the deterministic annealing approach to the minimization of $M(f \mid f^n; T)$ at each EM iteration for a given temperature T. Due to the non-convex roughness penalty, the M-step objective function is difficult to minimize. Following the techniques employed by (67; 39; 40; 70), *Iterated Conditional Modes (ICM)*, a coordinate-wise descent technique is used for energy minimization (5). Since this only decreases the M-step objective function at each iteration, it is considered a *generalized EM* algorithm. At each voxel index by (i, j, k), the objective function is minimized with respect to $f_{i,j,k}$ keeping all each voxel variables fixed. The next voxel is chosen in raster scan order, and the method is repeated. After a complete sweep of the image space, the procedure is repeated until some convergence criteria are met. This method is slow, but it will always converge to a local minimum (75, page 228). This local minimum is then used as the initial estimate for the next temperature setting and another local minimum will be found at this new setting. By varying tht temperature, the global minimum may be obtained. Unlike gradient descent methods, coordinate-wise descent techniques do not need to compute step sizes. The closed form solution for the minimum of each voxel can be obtained by differentiating the objective function, setting that result to zero and solving for the variable of interest. Consequently,

$$-\frac{\partial M(f \mid \hat{f}; T)}{\partial f_{ijk}} = -\frac{\partial Q(f \mid \hat{f})}{\partial f_{ijk}} + \frac{\partial E_P(f \mid \hat{f}; T)}{\partial f_{ijk}} \tag{5.43}$$

where

$$-\frac{\partial Q(f \mid \hat{f})}{\partial f_{ijk}} = \sum_{t,\theta} A_{t,\theta;i,j,k} - \sum d_{t,\theta} \frac{A_{t,\theta;i,j,k}\hat{f}_{i,j,k}}{\sum_{k,l} A_{t,\theta;m,n,o}\hat{f}_{m,n,o}} \frac{1}{f_{i,j,k}}. \tag{5.44}$$

In (68; 70), the line processes $\{f_{i-1,j}, f_{i,j-1}, f_{i,j}, f_{i,j+1}, f_{i,j-1}, f_{i+1,j+1}\}$, correspond to the nodes which are defined by the finite difference equations in Equations 3.10 to 3.19. Each of the line processes is evaluated to determine which of them will contribute to the objective function whose partial derivative will be minimized to form the update equation for a generic voxel f_{ijk}, i.e., which line processes produce nonzero quantities when applied to the objective function. For a three-dimensional lattice, the central voxel, its eight neighbors in the transverse plane and five neighbors in the adjacent transverse planes whose distance from the central voxel $f_{i,j,k}$ is less than two are used:

$$\begin{aligned} &\{l_{i,j,k}, l_{i+1,j,k}, l_{i-1,j,k}, l_{i,j-1,k}, l_{i,j+1,k}, l_{i,j,k-1}, \\ &l_{i,j,k+1}, l_{i+1,j-1,k}, l_{i-1,j-1,k}, l_{i-1,j+1,k}, l_{i+1,j+1,k}, l_{i+1,j,k-1}, \\ &l_{i-1,j,k-1}, l_{i,j-1,k-1}, l_{i,j+1,k-1}, l_{i+1,j,k+1}, l_{i-1,j,k+1}, l_{i,j-1,k+1}, l_{i,j+1,k+1}\} \end{aligned} \tag{5.45}$$

From the set of finite difference equations, only ten of the line processes produce nonzero value to be used in the GEM update equation:

$$\begin{aligned} &\{l_{i,j,k}, l_{i+1,j,k}, l_{i-1,j,k}, l_{i,j-1,k}, l_{i,j+1,k}, \\ &l_{i,j,k-1}, l_{i,j,k+1}, l_{i-1,j-1,k}, l_{i-1,j,k-1}, l_{i,j-1,k-1}\} \end{aligned} \tag{5.46}$$

Given these line processes, E_P can now be evaluated.

$$
\begin{aligned}
&\frac{\partial E_P(f \mid \hat{f}; T)}{\partial f_{ijk}} \\
&\qquad = \frac{1}{1+\exp\left[T\left(\beta U_{i,j,k}\right)\right]}\left\{36f_{i,j,k} - 12f_{i+1,j,k} - 12f_{i,j+1,k} - 12f_{i,j,k+1}\right. \\
&\qquad \left. -4f_{i-1,j,k} - 4f_{i,j-1,k} - 4f_{i,j,k-1} + 4f_{i+1,j+1,k} + 4f_{i+1,j,k+1} + 4f_{i,j+1,k+1}\right\} \\
&\qquad + \frac{1}{1+\exp\left[T\left(\beta U_{i+1,j,k}\right)\right]}\left\{2f_{i,j,k} + 2f_{i+2,j,k} - 4f_{i+1,j,k}\right\} \\
&\qquad + \frac{1}{1+\exp\left[T\left(\beta U_{i,j+1,k}\right)\right]}\left\{2f_{i,j,k} + 2f_{i,j+2,k} - 4f_{i,j+1,k}\right\} \\
&\qquad + \frac{1}{1+\exp\left[T\left(\beta U_{i,j,k+1}\right)\right]}\left\{2f_{i,j,k} + 2f_{i,j,k+2} - 4f_{i,j,k+1}\right\} \\
&\qquad + \frac{1}{1+\exp\left[T\left(\beta U_{i-1,j,k}\right)\right]}\left\{10f_{i,j,k} - 12f_{i-1,j,k} + 4f_{i-1,j+1,k} + 4f_{i-1,j,k+1}\right. \\
&\qquad \left. -4_{i,j,k+1} - 4f_i i, j+1, k+2f_{i-2,j,k}\right\} \\
&\qquad + \frac{1}{1+\exp\left[T\left(\beta U_{i,j-1,k}\right)\right]}\left\{10f_{i,j,k} - 12f_{i,j-1,k} + 4f_{i+1,j-1,k} + 4f_{i,j-1,k+1}\right. \\
&\qquad \left. -4f_{i+1,j,k} - 4f_{i,j,k+1} + 2f_{i,j-2,k}\right\} \\
&\qquad + \frac{1}{1+\exp\left[T\left(\beta U_{i,j,k-1}\right)\right]}\left\{10f_{i,j,k} + -12f_{i,j,k-1} + 4f_{i+1,j,k-1} + 4f_{i,j+1,k-1}\right. \\
&\qquad \left. -4f_{i+1,j,k} - 4f_{i,j+1,k} + 2f_{i,j,k-2}\right\} \\
&\qquad + \frac{1}{1+\exp\left[T\left(\beta U_{i-1,j,k-1}\right)\right]}\left\{2f_{i,j,k} + 4f_{i-1,j,k-1} - 4f_{i-1,j,k} - 4f_{i,j,k-1}\right\} \\
&\qquad + \frac{1}{1+\exp\left[T\left(\beta U_{i,j-1,k-1}\right)\right]}\left\{2f_{i,j,k} + 4f_{i,j-1,k-1} - 4f_{i,j-1,k} - 4f_{i,j,k-1}\right\} \\
&\qquad + \frac{1}{1+\exp\left[T\left(\beta U_{i-1,j-1,k}\right)\right]}\left\{2f_{i,j,k} + 4f_{i-1,j-1,k} - 4f_{i-1,j,k} - 4f_{i,j-1,k}\right\}
\end{aligned}
\tag{5.47}
$$

The derivation for $\frac{\partial E_P(f|\hat{f};T)}{\partial f_{ijk}}$ is given in Appendix B and follows the approach of Lee (68, pp. 107-108), who derived the objective function for the weak plate. It is based on the nonconforming finite element which was introduced in Section 3.1. The goal is to derive a closed-form solution for f from $\partial M(f \mid f^k; T) = 0$ at each voxel, while keeping all other voxel values fixed. It is obvious, however, that such an explicit solution is impossible due to the transcendental nature of the exponential terms. One possible

solution to the problem, which has been demonstrated in (67; 39; 40; 68; 70), is to separate the global dependence of the variables by introducing a new term z which is dependent upon f and descend upon a pair of variables f_{ijk} and z_{ijk} separately.

By examining the mean value of the line process variable l_i, the exponential term is equivalent to one minus the conditional mean value of l_i, suggesting that z_i is the conditional mean value of the line process l_i. Lee (68, pp. 109) identified z_{ij} from the mean values of the line processes conditioned on f and d at each T of the deterministic scheme by the following derivation:

$$\begin{aligned} E(l_{ij} \mid f, d; T) &= 1 \cdot P_T(l_{ij} = 1 \mid f, d) + 0 \cdot P_T(l_{ij} = 0 \mid f, d) \\ &= P_T(l_{ij} = 1 \mid f) \frac{P_T(f, l_{ij} = 1)}{P_T(f)} \\ &= \frac{\exp\{-T\beta\alpha\}}{\exp\{-T\beta\alpha\} + \exp\{-T\beta U(f_{ijk})\}} \\ &= \frac{1}{1 + \exp\{-T\beta(U(f_{ijk}) - \alpha)\}} \\ &= z_{ijk} \end{aligned} \tag{5.48}$$

From the definition of z_{ijk},

$$\begin{aligned} 1 - z_{ijk} &= \frac{\exp\{-T\beta(f_{ijk})\}}{\exp\{-T\beta\alpha\} + \exp\{-T\beta U(f_{ijk})\}} \\ &= \frac{1}{1 + \exp\{-T\beta\left[\alpha - U(f_{ijk})\right]\}} \end{aligned} \tag{5.49}$$

The objective function can now be expressed in terms of the new variables.

$$\begin{aligned} M(f, z \mid \hat{f}, T) = &\sum_{i,j,k} \sum_{t,\theta} \left[(A_{t\theta,ijk} f_{ijk} - d_{t\theta}) \frac{A_{t\theta;ijk} \hat{f}^n_{ijk}}{\sum_{klm} A_{t\theta;klm} \hat{f}^n_{klm}} \log\left(f_{ijk}\right) \right] \\ &+ \sum_{i,j,k} \left[\beta U_{i,j,k}(1 - z_{i,j,k}) + \alpha z_{i,j,k}\right] \\ &+ \frac{1}{T} \sum_{i,j,k} \left[z_{i,j,k} \log(z_{i,j,k}) + (1 - z_{i,j,k}) \log(1 - z_{i,j,k})\right]. \end{aligned} \tag{5.50}$$

The ICM update equations for $f_{i,j,k}$ are obtained by taking the partial derivative of $f_{i,j,k}$ with respect to Equation 5.50 and setting it to zero. The result can be rewritten as a quadratic equation which whose solution is the positive root computed by the quadratic formula. In the same manner, the update equation for $z_{i,j,k}$ can be derived by taking the partial derivative of Equation 5.50 with respect to $z_{i,j,k}$ and setting it to zero. Lee *et al* (67) shows that this objective function can be derived using Lagrange parameters.

Several researchers, including Leahy and Yan (65) and Gindi *et al* (40; 39; 67; 68; 70) derive GEM update equations by formulating the problem as a quadratic equation and using the quadratic formula to obtain a modified update equation

$$f_{i,j,k} = \frac{-(\sum_{t\theta} \mathcal{A}_{t\theta;i,j,k} - 2\lambda X_3) + \sqrt{(\sum_{t\theta} \mathcal{A}_{t\theta;i,j,k} - 2\lambda X_3)^2 + 8\lambda X_2 X_1}}{4\lambda X_2} \tag{5.51}$$

where X_1, X_2 and X_3 are defined for weak membranes in (67; 40; 39) and weak plates in (68; 70). In three dimensions, the weak membrane may be extended trivially using the update equation with X_2 and X_3 defined as

$$\begin{aligned} X_2 &= (1 - z^h_{i,j,k}) + (1 - z^h_{i-1,j,k}) + (1 - z^v_{i,j,k}) \\ &\quad + (1 - z^h_{i,j-1,k}) + (1 - z^t_{i,j,k}) + (1 - z^t_{i,j,k-1}) \\ X_3 &= f_{i+1,j,k}(1 - z^h_{i,j,k}) + f_{i-1,j,k}(1 - z^h_{i-1,j,k}) + f_{i,j+1,k}(1 - z^v_{i,j,k}) \\ &\quad + f_{i,j-1,k}(1 - z^h_{i,j-1,k}) + f_{i,j,k+1}(1 - z^t_{i,j,k}) + f_{i,j,k-1}(1 - z^t_{i,j,k-1}) \end{aligned} \tag{5.52}$$

As derived in Appendix B, extending the weak plate in three dimensions yields the update equations

$$
\begin{aligned}
X_2 &= 21(1 - z_{i,j,k}) + 5(1 - z_{i-1,j,k}) + 5(1 - z_{i,j-1,k}) \\
&\quad + 5(1 - z_{i,j,k-1}) + (1 - z_{i+1,j,k}) + (1 - z_{i,j+1,k}) \\
&\quad + (1 - z_{i,j,k+1}) + (1 - z_{i-1,j-1,k}) + (1 - z_{i-1,j,k-1}) \\
&\quad + (1 - z_{i,j-1,k-1}) \\
X_3 &= \{6\,(f_{i+1,j,k} + f_{i,j+1,k} + f_{i,j,k+1}) \\
&\quad + 2\,(f_{i-1,j,k} + f_{i,j-1,k} + f_{i,j,k-1}) \\
&\quad - 2\,(f_{i+1,j+1,k} + f_{i,j+1,k} + f_{i+1,j,k+1})\}\,(1 - z_{i,j,k}) \\
&\quad + \{6f_{i-1,j,k} + 2\,(f_{i+1,j-1,k} + f_{i,j-1,k+1}) \\
&\quad - 2\,(f_{i+1,j+1,k} + f_{i-1,j,k+1}) + f_{i+2,j,k}\}\,(1 - z_{i-1,j,k}) \\
&\quad + \{6f_{i,j-1,k} + 2\,(f_{i+1,j,k} + f_{i,j,k+1}) \\
&\quad - 2\,(f_{i,j-1,k} + f_{i+1,j-1,k}) + f_{i,j-2,k}\}\,(1 - z_{i,j-1,k}) \\
&\quad + \{6f_{i,j,k-1} + 2\,(f_{i+1,j,k} + f_{i,j,k+1}) \\
&\quad - 2\,(f_{i,j-1,k} + f_{i+1,j-1,k}) + f_{i,j,k-2}\}\,(1 - z_{i,j,k-1}) \\
&\quad + \{2f_{i+1,j,k} - f_{i+2,j,k}\}\,(1 - z_{i+1,j,k}) \\
&\quad + \{2f_{i,j+1,k} - f_{i,j+2,k}\}\,(1 - z_{i,j+1,k}) \\
&\quad + \{2f_{i,j,k+1} - f_{i,j,k+2}\}\,(1 - z_{i,j,k+1}) \\
&\quad + \{2\,(f_{i-1,j,k} + f_{i,j-1,k}) - 2f_{i-1,j-1,k}\}\,(1 - z_{i-1,j-1,k}) \\
&\quad + \{2\,(f_{i-1,j,k} + f_{i,j,k-1}) - 2f_{i-1,j,k-1}\}\,(1 - z_{i-1,j,k-1}) \\
&\quad + \{2\,(f_{i,j-1,k} + f_{i,j,k-1}) - 2_f i, j-1, k-1\}\,(1 - z_{i,j-1,k-1})
\end{aligned}
\tag{5.53}
$$

When all of the line processes are zero, the update equations revert to the biharmonic equation defined in Section 3.2. With these update equations, the deterministic annealing algorithm can be defined:

Data: projection data d, $\alpha_0, \beta_0, \epsilon_0$, L
Result: Reconstructed volume V
begin
  $\alpha \longleftarrow \alpha_0$;
  $\beta \longleftarrow \beta_0$;
  $\epsilon \longleftarrow \epsilon_0$;
  $z_{ijk} \longleftarrow l_{ijk} \forall\, ijk$;
  while $\exists\ z_{ijk} > \tau$ *or* $z_{ijk} \leq (1-\tau)$**do**
    while *MRF energy change* $> \epsilon$*;*
    do
      $f^{(k+1)} = PML\text{-}GEM\ update\ ;$
      $z^{(k+1)} = (1-\lambda)\frac{1}{1+exp(-\beta T(\Delta-\alpha))} + \lambda z^{(k)}$*;*
    end
    $T \longleftarrow T/2$*;*
    $\epsilon \longleftarrow \epsilon/2$*;*
  end
end

Algorithm 2: Deterministic Annealing

Lee *et al* (68; 70) employs an annealing schedule that uses a doubling scheme for T at each epoch. Iterations are performed at a given T until the relative energy change $\frac{E_n - E_{n-1}}{E_n - E_0} < \epsilon$, where n indexes the iteration number within a epoch. The entire simulation is terminated when $z_{ijk} \leq \tau$ or $z_{ijk} \geq 1 - \tau$, where τ is chosen as a threshold. Since z corresponds to the conditional expected value of the line process l, the approach of z to zero or one corresponds to the approach of the ϕ function to the broken parabola.

Arbitrarily changing variables in the objective function and then performing cyclic coordinate-wise descent on the new variables does not guarantee convergence unless there exists a corresponding objective function that can be expressed in terms of the new variables. Appendices A and B demonstrate that such a function can be derived using Lagrange parameters.

With Algorithm 2 providing a framework for using deterministic annealing with 3-D thin plate splines as a smoothing mechanism, Algorithm 3 can be presented as an approach for reconstructing dual-head SPECT imaging using PML-GEM.

Data: System matrices $A_{0.5}$ and $A_{3.0}$, Projection data sets $d_{0.5}$ and $d_{3.0}$
Result: Reconstructed volume V
begin
    Perform dual-head SPECT imaging with different collimators
        Acquire high-resolution data from 0.5 mm collimator
        Acquire high-sensitivity data from 3.0 mm collimator
    Reconstruct high-resolution data using MLEM
    Extract edge profiles L from high-resolution data
    Reconstruct high-sensitivity data using Algorithm 2
end

Algorithm 3: Dual-head SPECT reconstruction

The initial line process estimates are obtained from the initial reconstruction using sigmoidal thresholding. The parameters α, β and ϵ are established for each reconstruction. Conversely, the initial reconstruction could be from X-ray transmission tomography or anatomical priors instead of a SPECT reconstruction. If memory is sufficient, the second reconstruction could use the data sets simultaneously.

Chapter 6

Computational Experiments and Results

Computational experiments have been conducted using mathematical phantoms and data from Monte Carlo simulations of the Siemens Multimodality Inveon scanner. A single volumetric data set of a spleen has been acquired through X-ray computed tomography of a spleen and has been segmented by a human observer to obtain a mask that can indicate the position of the corresponding tissue in the emission tomography. Due to resolution mismatch between the X-ray CT and SPECT imaging modalities, the mask must be downsampled and blurred to match the dimensions of the reconstructed SPECT data. The resulting mask can then be used to calculate the relative count and relative percentage of count recovery which are among the criteria using to evaluate the efficacy of Bayesian reconstruction techniques. As the downsampled mask may not accurately reflect the blur introduced by the optical transfer function of the SPECT system, it may need to be either dilated using morphological operators or blurred using Gaussian kernels.

Once the mask has been created, SPECT data is reconstructed using MLEM, MAP-GEM or PML-GEM algorithms. The penalty functions used for MAP-GEM

is the quadratic prior and for PML-GEM is the thin hyperplate spline defined in Chapter 5.

6.1 Methodology

In the simulations, attenuation, scatter and detector response have been excluded from consideration, but these can be incorporated at a later time. The forward and back projector operators have been modeled following the approach of Gregor *et al* (44). For a select number of geometric shapes, projection data are generated for a 360 degree spherical orbit traversed by 60 equally-spaced projection angles. A two-dimensional detector with 68×68 pixels has been modeled. The projection data has been oversampled by a factor of two in an attempt to mitigate any ringing artifacts that are present due to ripples in the sinc function which is the impulse response of the perfect low-pass filter used to generate the data. Mathematically, this ripple effect is called the *Gibbs phenomenon* and refers to the behavior of a Fourier series of a piecewise continuously differentiable periodic function at a sharp discontinuity, such as an edge in an image (86).

The mathematical phantoms are of a sphere with uniform activity distribution, centered in the scanner's field of view and that same phantom with hot and cold regions of interest. Additional data has been generated from a mouse spleen that was segmented manually from an X-ray computed tomography scan (2). A constant pixel value of 100 was assigned to each of the voxels in the uniform phantom and the projection data was generated. A hot lesion with twice the uptake of the organ was modeled as was a cold region which contains half the uptake value of the organ itself. Volumetric renderings of the mathematical phantoms are shown in Figure 6.2 and the statistics for each region of interest are presented in Tables 6.1 and 6.2. The uniform sphere phantom and the hot and cold phantom serve different purposes in analyzing the effectiveness of different penalty functions. The uniform sphere phantom provides an ideal subject for determining the spatial resolution for various

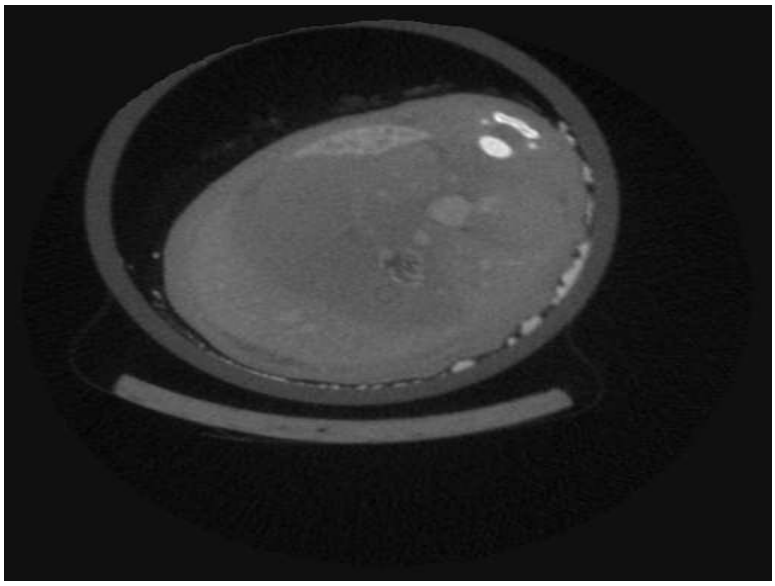


Figure 6.1: X-ray CT slice of a mouse

reconstruction algorithms while the hot and cold phantom provides an ideal phantom for calculating contrast sensitivity at different signal-to-noise ratios. As the small size and irregular shape of the mouse spleen poses a challenge for traditional analysis, a count recovery mechanism is employed to assess the performance of reconstruction algorithms.

6.2 Experiments with Mathematical Phantoms

The mathematical phantoms have been generated by forward projection of geometric shapes through the system probability matrix. A Poisson distribution with the mean set to the projected data value is used to generate the simulated data for the reconstruction experiments. For each projection angle, a new Poisson estimate is realized and projected through the system matrix to generate a projection image. Data sets were generated with one or more region of interests and a background mean voxel value of zero. After the image had been forward projected through two probability matrices, one with a simulated pinhole collimator diameter of 0.5 mm and the other with a pinhole diameter of 3.0 mm, the data was generated and Gaussian

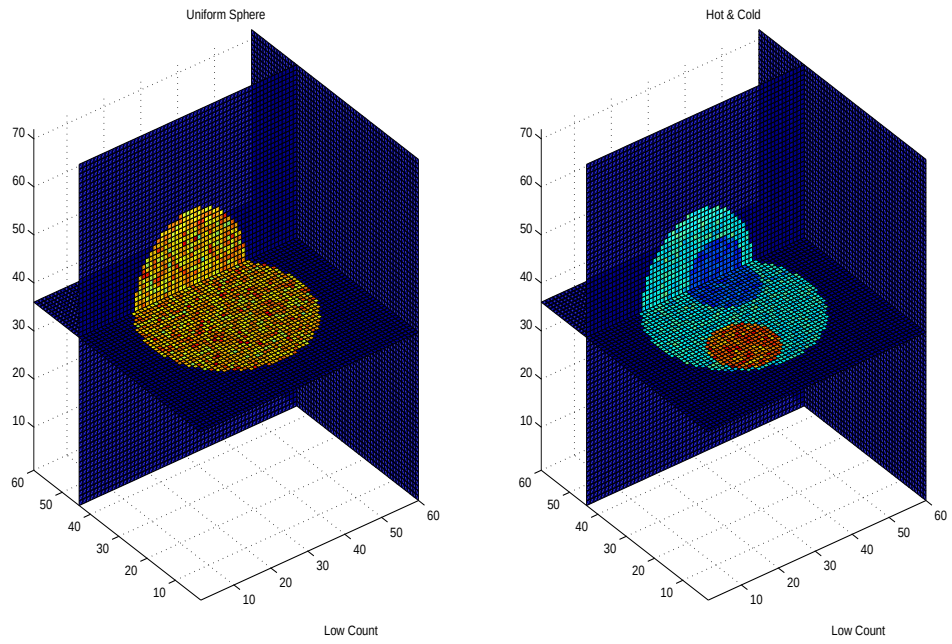


Figure 6.2: Mathematical Phantoms

noise was injected into the system to simulate detector noise. Neither scatter nor attenuation effects were simulated.

For the uniform sphere, a single region of interest comprises the entire sphere, which has a radius of 20 mm. For the hot and cold phantom, there are three regions of interest, include one hot region with a radius of 8 mm and a cold region of interest with a radius of 8 mm. Both hot and cold regions are centered in the transaxial plane.

Tables 6.1 and 6.2 provide the statistics for the spherical phantom and each of the regions of interest in the hot and cold spherical phantom, respectively.

To simulate actual clinical data, a large number of simulations should be generated that use Poisson noise for the spherical phantom. The mathematical phantoms were

Table 6.1: ROI Statistics in the ideal mathematical phantom

Region of Interest	Number of Voxels	True Value	Mean	Standard Deviation
Background	259,200	0	0	0
Organ	33,552	100	100.14	10.14

Table 6.2: ROI Statistics in the ideal spherical phantom with hot and cold spots

Region of Interest	Number of Voxels	True Value	Mean	Standard Deviation
Background	259,200	0	0	0
Organ	29,200	100	99.91	9.96
Hot lesion	2,176	200	199.75	13.79
Cold spot	2,176	50	49.88	6.98

corrupted with 20 dB, 30 dB and 50 dB of average white Gaussian noise (awgn), respectively. Figure 6.3 shows a projection image at each noise level. The noise level can have a dramatic impact on the image quality of the reconstructed volume.

Once the mathematical phantoms have been generated, a series of ML-EM and MAP-GEM reconstructions are performed to determine the effect of the smoothing penalties. Following the approach used by Lee *et al*, (70), the χ^2 stopping rule of Llacer and Veklerov (74) was used to terminate the reconstructions. The hyperparameters were chosen empirically.

6.3 Experiments with Monte Carlo simulations

Monte Carlo simulations with the GATE software (53) generated a set of data sets for reconstruction. Balb-C wild type mouse data sets were obtained, spleens were manually segmented and a single data set was used as the basis for GATE simulations. From the original emission map containing $512 \times 512 \times 768$ voxels, a decimated spleen data set served as a model for the GATE simulations which produced two mouse spleen data sets. Sample projection images are shown in Figure 6.5 and a volume rendering of the reconstructed mouse spleen is shown in Figure 6.6.

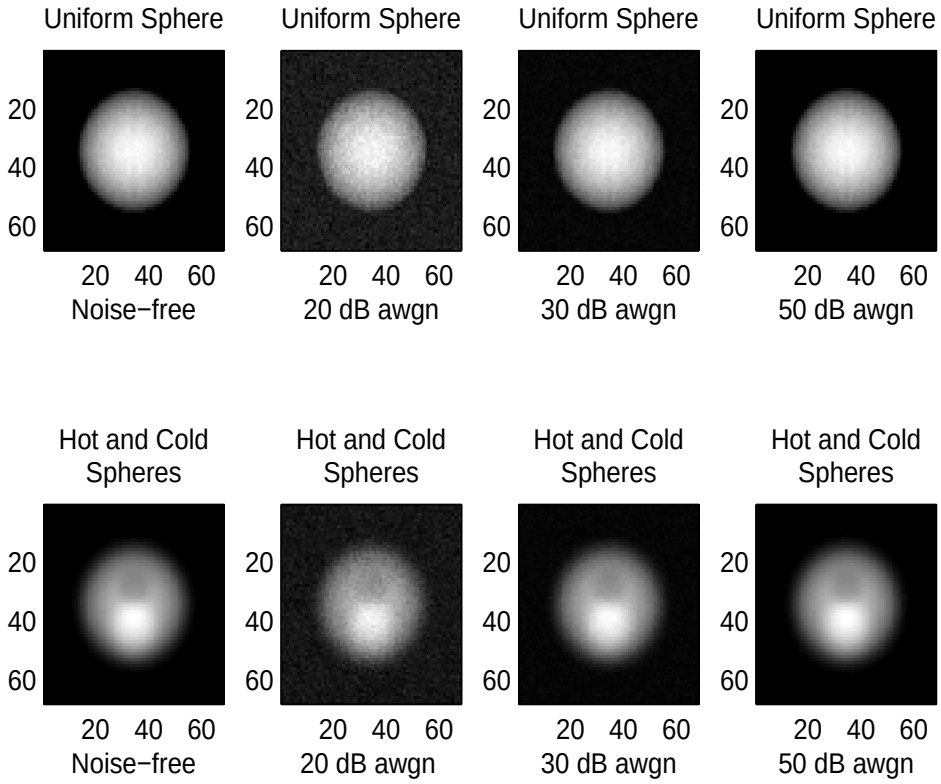


Figure 6.3: Noisy Projection Images at 20, 30 and 50 dB

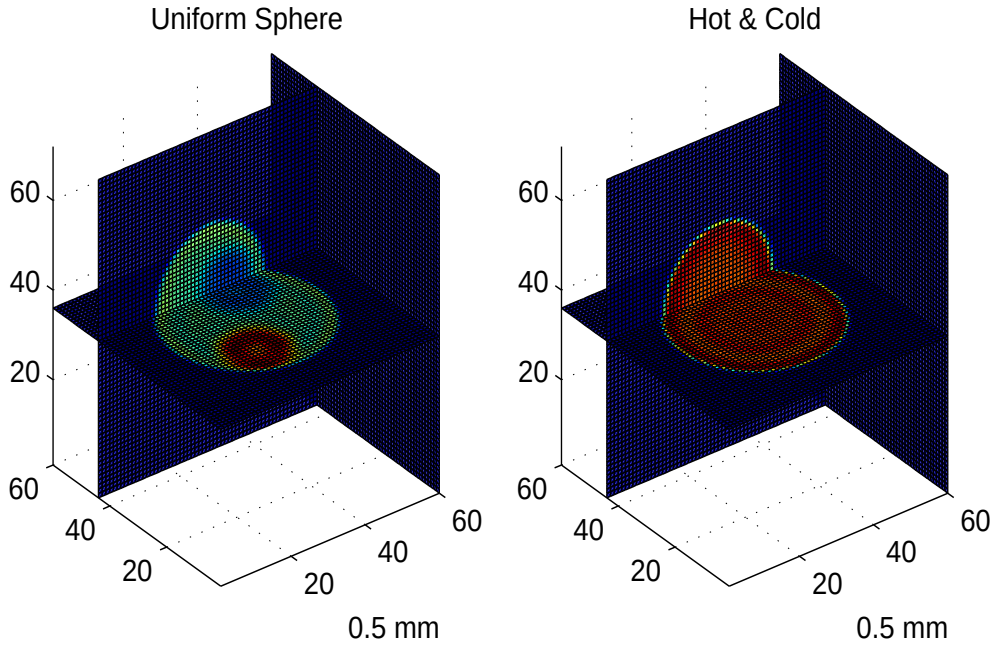


Figure 6.4: Reconstructed Volumes

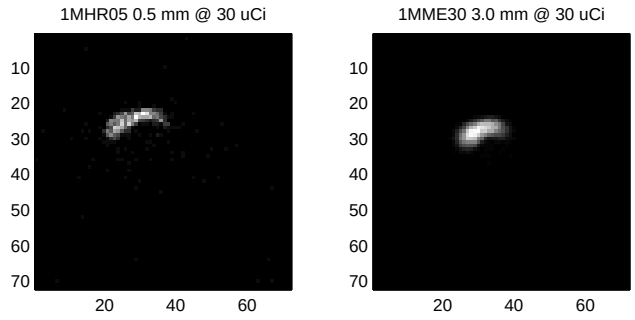


Figure 6.5: GATE-generated mouse spleen data

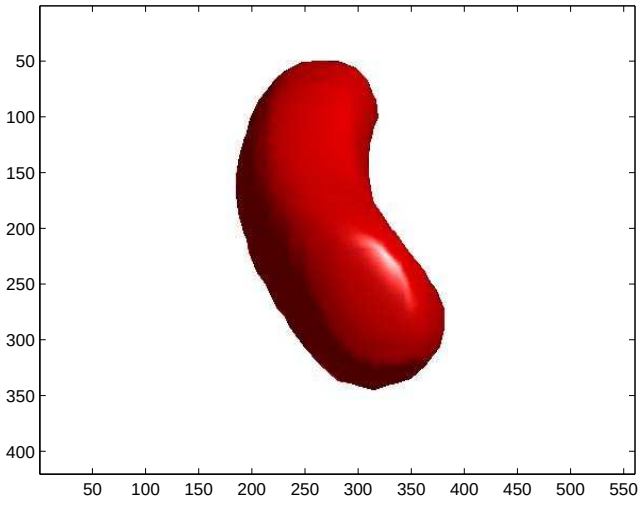


Figure 6.6: Reconstructed spleen volume visualization

To assess the quality of reconstructed images, Carson *et al* (15) and Lee *et al* (70) computed bias and standard deviation images for a large collection of reconstructions. A bias image, b_{ij}, is defined as

$$b_{ij} \stackrel{\text{def}}{=} \frac{1}{K} \sum_{k=1}^{K} (\hat{f}_{ij}^{k} - f_{ij}) \tag{6.1}$$

where f_{ij} is the phantom's voxel value at location (i, j, k) and $\hat{f}_{ij}^{k}$ is the kth reconstruction of phantom f at location (i, j) and the summation is over K independent noise trials. A standard deviation image, σ_{ij}, is defined as

$$\sigma_{ij} \stackrel{\text{def}}{=} \sqrt{\frac{1}{(K-1)} \sum_{k=1}^{K} (\hat{f}_{ij}^{k} - \bar{f}_{ij})^2} \tag{6.2}$$

where $\bar{f}_{ij}$ is the mean of $\hat{f}_{ij}$ over the noise trials defined as $\bar{f}_{ij} = (1/K) \sum_{k=1}^{K} \hat{f}_{ij}^{k}$. The same normalization is used for all images in a given figure to permit comparison. Multiple regions of interest were chosen randomly and the bias and standard deviation were calculated for each region. Each line segment were centered at the bias with the spread being equal to a single standard deviation. Both bias and standard deviation were expressed as a percentage of the corresponding true regional mean. This methodology could not be followed by this research as there were too few data sets to constitute a sufficient sample size for the bias and standard deviation calculations. Future research may provide a sufficient number of data sets to utilize these techniques for image quality assessment.

6.4 Modulation Transfer Function

The modulation transfer function (MTF) is the most widely used resolution analysis for imaging systems. It describes the reduction in contrast that occurs when a spatial sinusoid is captured by an imaging system and is defined as the ratio of the contrast of

the measured sinusoidal output to the contrast of the sinusoidal input. In practice, a spatial sinusoid is not the image being acquired and the MTF is defined as the discrete Fourier transform of the line spread function (LSF). The LSF can be calculated by two different methods. It can be found directly from an ideal line approximation provided by a slit test target or it can be derived from the edge spread function (ESF). In evaluating the ESF, an operator defines an area which encompasses the edge of the target image. The image pixel data is translated into a two-dimensional array (pixel intensity and pixel position). The amplitude (pixel intensity) of each line within the array is normalized and averaged, yielding the ESF. The LSF is then calculated as the first derivative of the ESF, which is obtained using numerical techniques such as a finite difference method. The MTF is then plotted against spatial frequency.

Figure 6.7 shows the MTF curves for reconstruction algorithms using the 0.5 mm collimator. In this experiment, 0.5 mm projection data is reconstructed with the MLEM algorithm, MAP-GEM with a quadratic prior and PML-GEM with the thin plate spline prior. As expected, MLEM performs best, but the PML-GEM algorithm performed better than MAP-GEM with a quadratic prior. In the next experiment, 3.0 mm projection is reconstructed with the MLEM algorithm, MAP-GEM with a quadratic prior and PML-GEM with the thin plate spline prior. The thin plate prior with a penalty based on 0.5 mm projection data performs better than both MLEM and MAP-GEM with a quadratic prior. Figure 6.8 shows the MTF curves for reconstruction algorithms using a 3.0 mm collimator.

For the Hot and Cold phantom, the regions of interest were segmented prior to computation of the MTF curves. Figures 6.9 and 6.10 show the MTF curves for the PML-GEM algorithm versus that of MLEM and MAP-GEM with a quadratic prior. In both the hot and cold regions, the PML-GEM algorithm performed better than the other two reconstruction algorithms using a 3.0 mm collimator. All of the algorithms performed better on the hot region of interest than the cold region due to the increased count level.

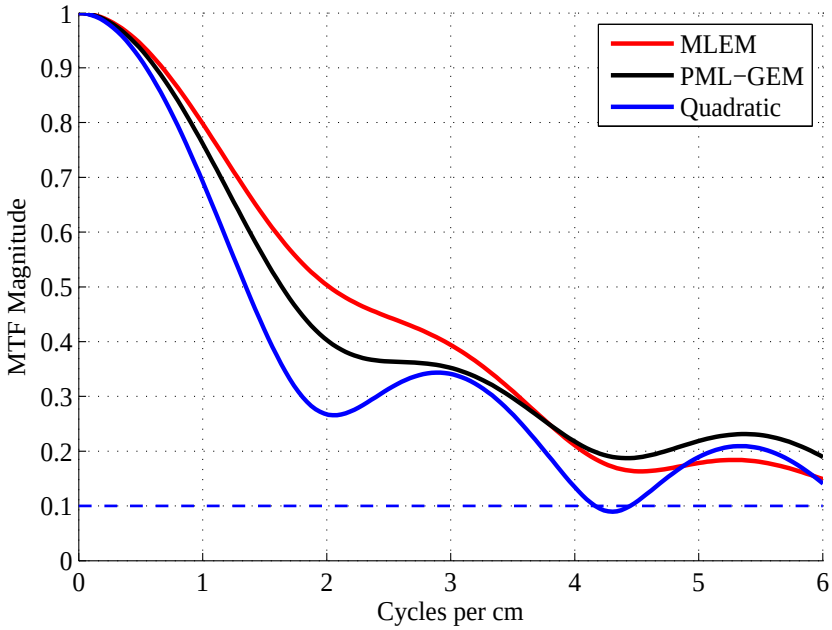


Figure 6.7: MTF curves for 0.5 mm pinhole collimator

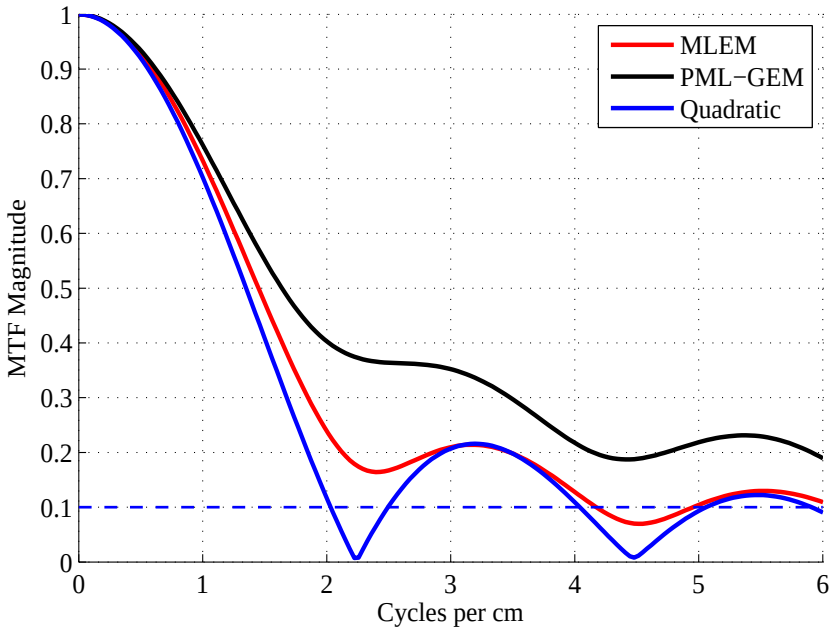


Figure 6.8: MTF curves for 3.0 mm pinhole collimator

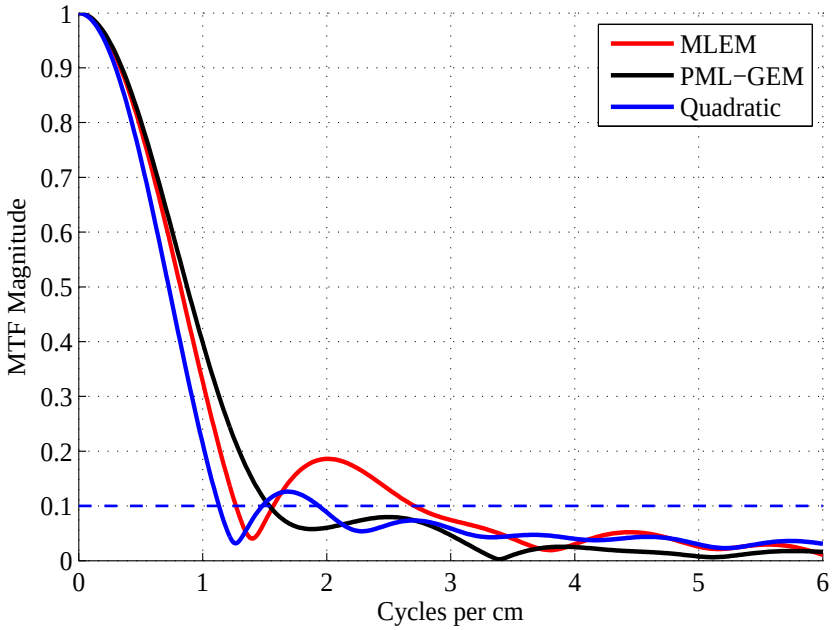


Figure 6.9: Cold MTF curves for 3.0 mm pinhole collimator

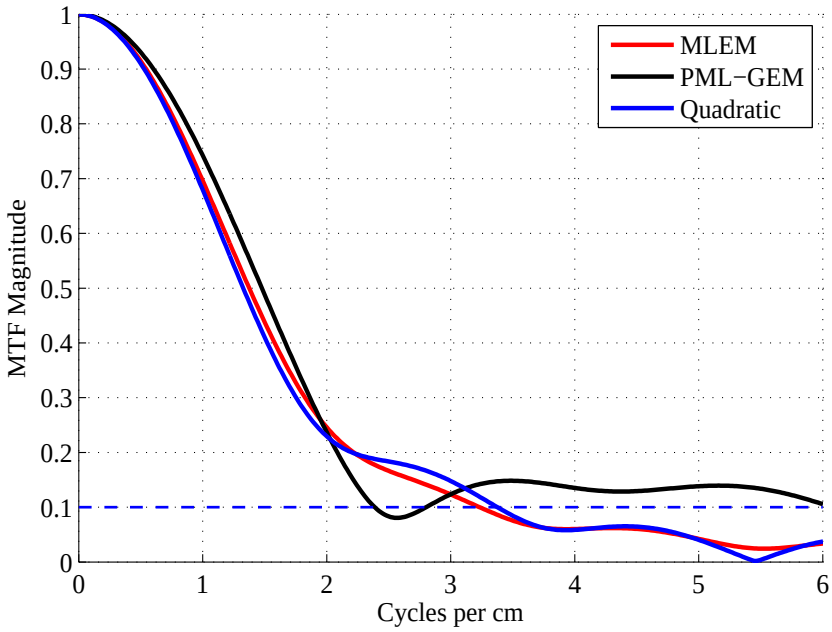


Figure 6.10: Hot MTF curves for 3.0 mm pinhole collimator

The PML-GEM algorithm with the thin plate penalty performed better than MLEM or MAP-GEM with the quadratic prior on reconstructions of the 3.0 mm single pinhole projection data.

6.5 Contrast Sensitivity

Contrast sensitivity is a measure of the contrast required in an image to discriminate an object from the background. The region of interest is typically a cylinder and must be uniform to obtain an accurate measurement. The region is subdivided into neighborhoods of size $N \times N \times N$, the mean value for each neighborhood is calculated and the standard deviation of the means is calculated. This process is repeated for $N = 1, 2, \ldots, M$ until there are fewer than 25 neighborhoods within the slice. The standard deviation is multiplied by three and plotted as a function of neighborhood size to produce the contrast discrimination function (CDF).

In the next set of experiments, the contrast discrimination function is used to evaluate the reconstruction algorithms against projection data which have been corrupted with 20 dB noise. Figure 6.11 shows the CDF curves for reconstruction algorithms using the 0.5 mm collimator. As before, 0.5 mm projection data is reconstructed with the MLEM algorithm, MAP-GEM with a quadratic prior and PML-GEM with the thin plate spline prior. The PML-GEM algorithm performed better than MAP-GEM with a quadratic prior or MLEM. The reconstructed volume exhibits the best statistics and retains this advantage as the region being averaged continues to expand. In the next experiment, 3.0 mm projection is reconstructed with the same reconstruction algorithms and the same results occur. PML-GEM produces a smoother reconstruction than either MLEM or MAP-GEM with a quadratic prior. Figure 6.12 shows the CDF curves for reconstruction algorithms using the 3.0 mm collimator.

For the Hot and Cold phantom, the regions of interest were segmented prior to computation to computation of the CDF curves. The reconstructed volumes were

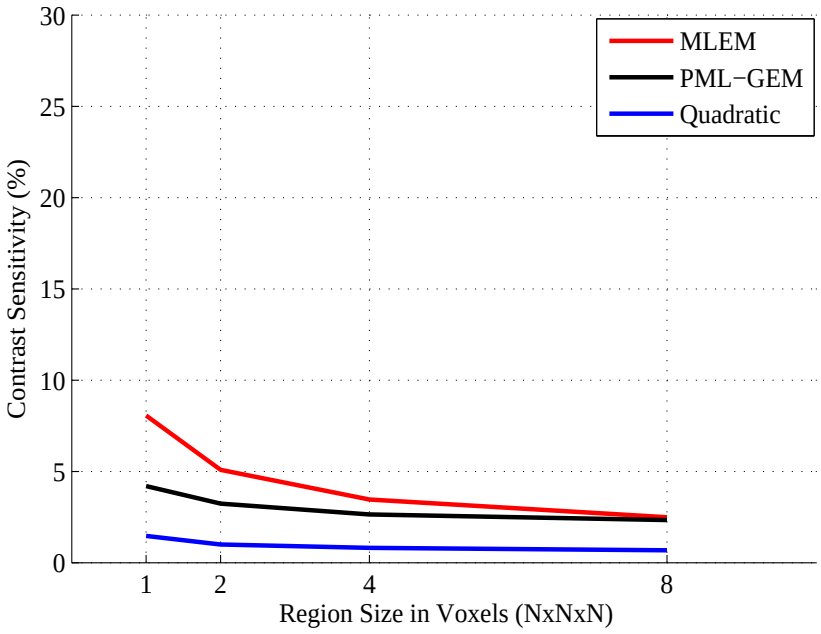


Figure 6.11: CDF curves for 0.5 mm pinhole collimator (20 dB)

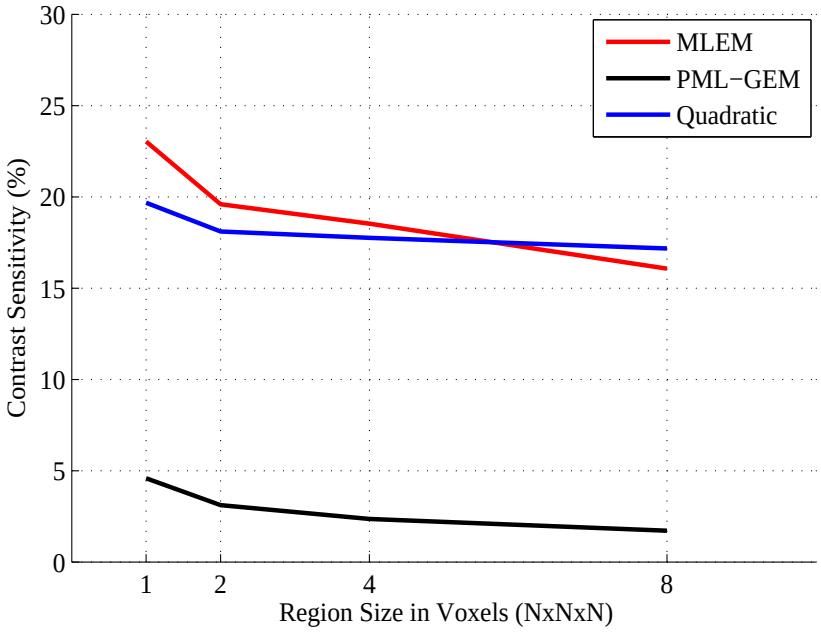


Figure 6.12: CDF curves for 3.0 mm pinhole collimator (20 dB)

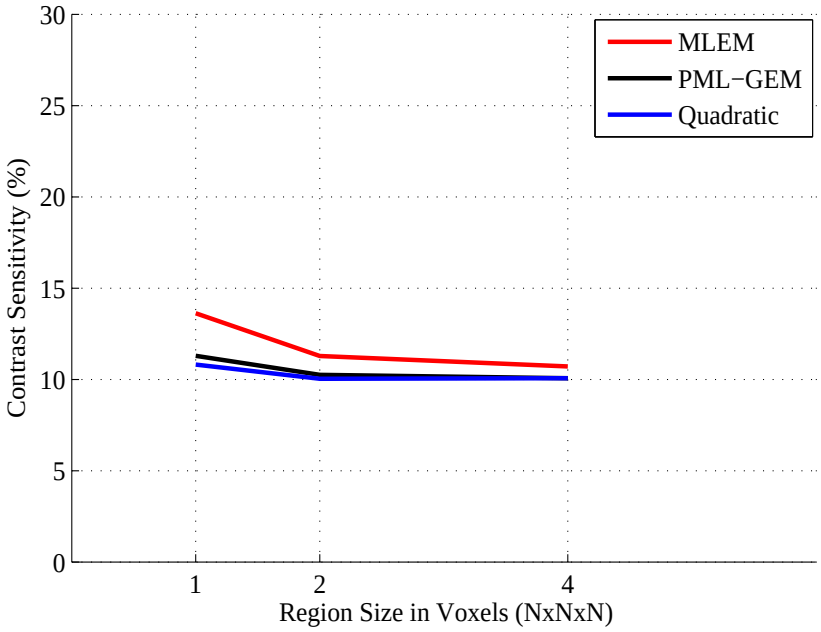


Figure 6.13: Cold CDF curves for 3.0 mm pinhole collimator (20 dB)

from the simulated 3.0 mm pinhole data with 20 dB awgn. Figures 6.13 and 6.14 show the CDF curves for the PML-GEM algorithm versus that of MLEM and MAP-GEM with a quadratic prior. In both the hot and cold regions, the PML-GEM algorithm performed as well as MAP-GEM with a quadratic prior. Combined with the better MTF curves, the PML-GEM algorithm with the thin spline prior was superior to the other reconstruction algorithms.

6.6 Count recovery metrics

When the region of interest is an irregular shape or is too small to be analyzed by conventional performance measures, an analysis of the counts recovered in the region of interest may serve as a measure of performance. A masked region is created, either from an anatomical prior or some threshold value, and all of the counts that region are tabulated and compared against all of the emitted counts. The masked region may be expanded using morphological operators.

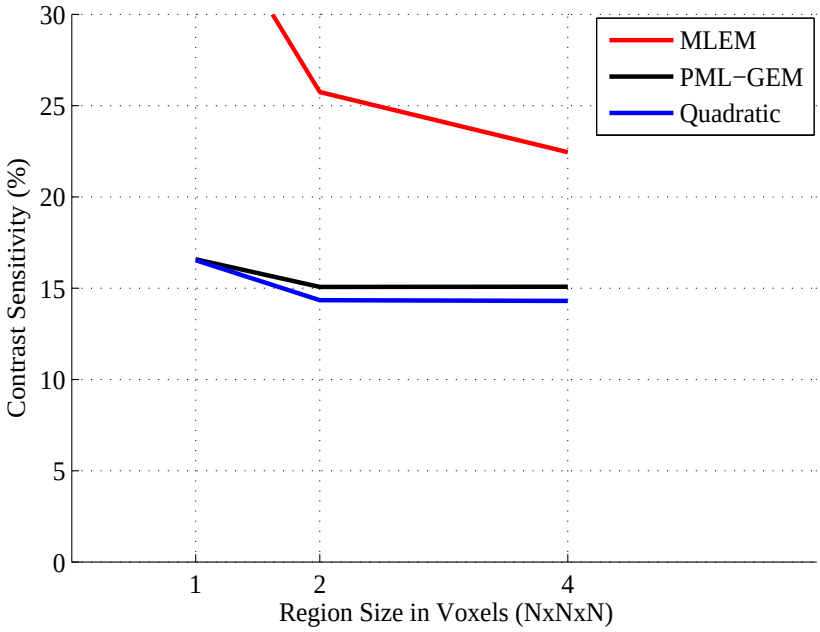


Figure 6.14: Hot CDF curves for 3.0 mm pinhole collimator (20 dB)

This experiment takes the mouse spleen data generated by GATE. In this experiment, a higher relative percentage of counts is reconstructed within the masked region by PML-GEM with the thin plate prior than the other algorithms. Figure 6.15 shows the relative count recovery of the three reconstruction algorithms for the mouse spleen data. The PML-GEM algorithm captures a larger percentage of relative counts at an earlier stage of the reconstruction.

While the PML-GEM algorithm with a thin hyperplate or biharmonic spline appears to successfully penalize high-sensitivity, lower-resolution single pinhole reconstructions, additional experiments need to be conducted with greater numbers of data sets with differing noise characteristics to validate the results using statistical measures. In addition, this approach to penalized image reconstruction has not been tested on multiple-pinhole collimators, which are becoming more prevalent in preclinical SPECT settings.

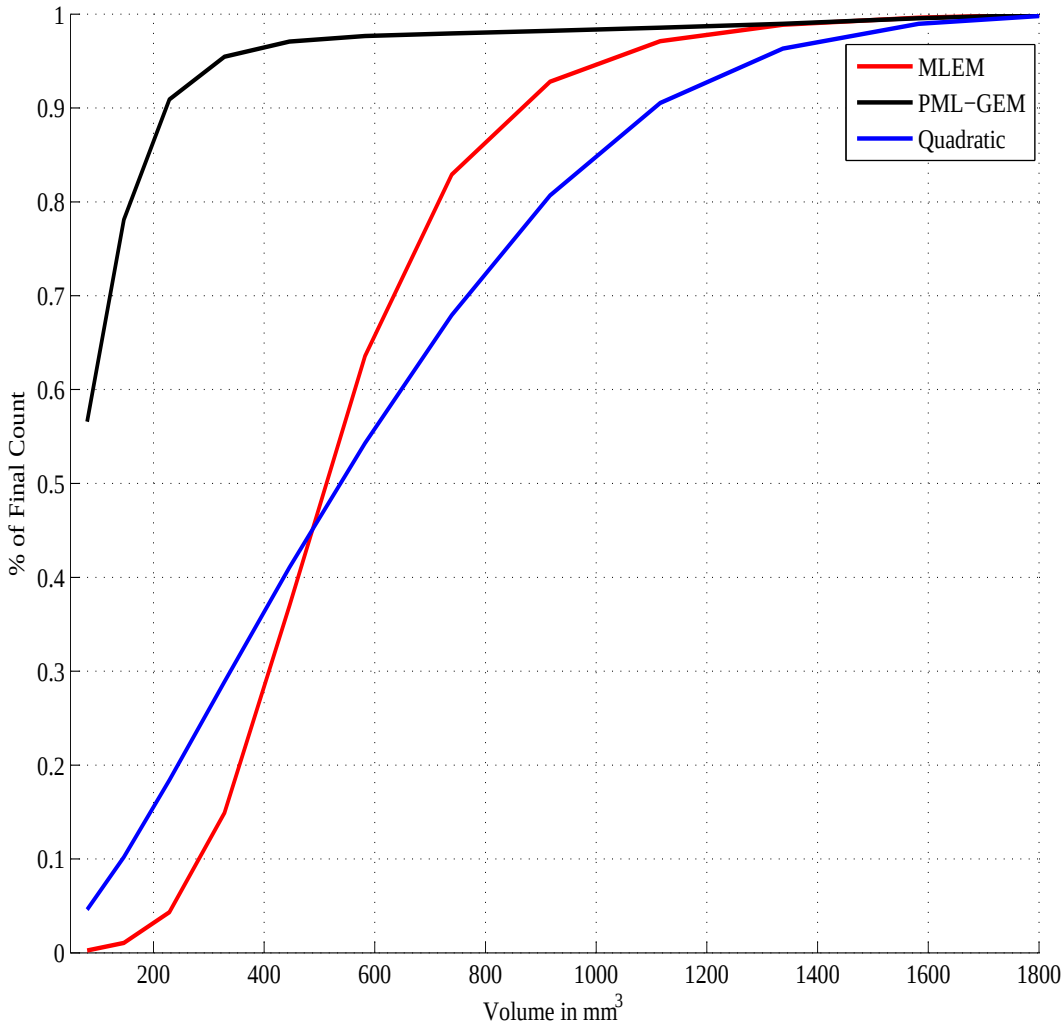


Figure 6.15: Relative Count Recovery

Chapter 7

Conclusions

The goal of this research is to investigate the use of prior data to improve estimation of the location and density of radiopharmaceutical uptake in living tissue, thereby improving diagnostic accuracy. Diagnostic accuracy benefits from better methods to quantify metabolic function within organs. A common approach to quantifying activity is to register anatomical and functional images, using an anatomical image as a guide to segmenting functional images for the purpose of identifying activity within organs or tissue. Partial volume effects may complicate the process of quantifying metabolic activity attributed to organs as boundaries extracted from anatomical images may not encompass all metabolic activity. More accurate reconstructions offer the promise of better quantification of radiopharmaceutical uptake.

7.1 Contributions

This research extends previous research in penalized image reconstruction by extending the mechanical models of (39; 40; 67; 70; 69) through the extension of a nonconforming finite element model proposed by (92) into three dimensions. It provides a new derivation of a mechanical prior for three-dimensional SPECT data and combines dual collimator SPECT projection data by using high-resolution data to penalize lower-resolution data. It compares the new three-dimensional

penalized reconstruction technique with existing penalized techniques through the use of modulation transfer function (MTF) and contrast transfer function (CTF) curves as prescribed in the ASTM International E1695 standard. Lastly, it suggests areas for future research, including the analysis of more complicated linear elasticity models as the basis for future penalized reconstruction techniques and the use of alternate sigmoid transfer functions in the formulation of line processes from reconstructed data.

7.2 Future research

The MAP-MRF framework can be applied in numerous imaging applications but the resulting energy minimization problems have been viewed as intractable. Recently, newer approximate energy minimization algorithms such as graph cuts, loopy belief propagation and tree-reweighted message passing have supplanted earlier approaches such as simulated annealing and Iterative Conditional Modes (ICM). Researchers have noted that these new approaches often use energy formulations targeted to the application (91). Woodford *et al* (99) demonstrate that inference with higher-order cliques can be performed using these newer techniques.

There has been a significant shift towards first-order smoothness priors as optimization techniques such as graph cuts and belief propagation networks have been developed. Boykov and Veksler (11) remark that it is unclear whether triple or higher-order cliques can be used to encode higher-order smoothness. Weak membrane models may be more conductive to graph cut techniques and may warrant additional research.

Future research might concentrate on obtaining preclinical data from a Siemens Inveon Multimodality scanner operating with single and multiple pinhole collimators. The mathematical phantoms and GATE simulations have generated data consist with that acquired from single pinhole collimators for the Inveon have pinhole diameters

of 0.5, 1.0, 2.0 and 3.0 mm, respectively. This research has concentrated upon using the smaller pinhole to constraint the reconstruction of the larger pinhole.

Bibliography

[1] M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions*, Dover Publications, Inc, New York, 1965. 29

[2] D. Aykac, J. R. Price, and J. S. Wall, *3D segmentation of the mouse spleen in microCT via active contours*, in Proceedings of the 2005 IEEE Nuclear Science Symposium and Medical Imaging Conference, 2005, pp. 1542–1545. 77

[3] L. Bedini, I. Gerace, and A. Tonazzini, *Sigmodial approximations for self-interacting line processes in edge-preserving image restoration*, Pattern Recognition Letters, 16 (1995), pp. 1011–1022. 56

[4] J. Besag, *Spatial interaction and the statistical analysis of lattice systems (with discussions)*, Journal of the Royal Statistical Society, Series B, 36 (1974), pp. 192–236. 31, 33, 35

[5] ——, *On the statistical analysis of dirty pictures (with discussions)*, Journal of the Royal Statistical Society, Series B, 48 (1986), pp. 259–302. 68

[6] M. Bierlaire, P. L. Toint, and D. Tuyttens, *On iterative algorithms for linear least squares problems with bound constraints*, Linear Algebra and its Applications, 143 (1991), pp. 111–143. 55

[7] M. Black and A. Rangarajan, *On the unification of line processes, outlier rejection, and robust statistics with applications in early vision*, International Journal of Computer Vision, 19 (1996), pp. 57–91. 49

[8] A. Blake and A. Zisserman, *Visual Reconstruction*, MIT Press, Cambridge, Massachusetts, 1987. 4, 6, 27, 36, 38, 41, 42, 44, 47, 49, 56, 57, 59, 66

[9] F. L. Bookstein, *Principal warps: Thin-plate splines and the decomposition of deformations*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 11 (1989), pp. 567–585. 29

[10] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, 2004. 39

[11] Y. Boykov and O. Veksler, *Graph cuts in vision and graphics: Theories and applications*, in Handbook of Mathematical Models in Computer Vision, N. Paragios, Y. Chen, and O. Faugeras, eds., Springer-Verlag, New York, 2006, ch. 5, pp. 79–111. 94

[12] J. Browne and A. R. De Pierro, *A row-action alternative to the EM algorithm for maximizing likelihoods in emission tomography*, IEEE Transactions on Medical Imaging, 15 (1996), pp. 687–699. 4

[13] T. F. Budinger and G. T. Gullberg, *Three-dimensional reconstruction in nuclear medicine emission imaging*, IEEE Transactions on Nuclear Science, 21 (1975), pp. 789–798. 10

[14] C. Byrne, *Applied Iterative Methods*, A. K. Peters, Ltd., Wellesley, Massachusetts, 2008. 2

[15] R. E. Carson, Y. Yan, B. Chodkowski, T. K. Yap, and M. E. Daube-Witherspoon, *Precision and accuracy of regional radioactivity quantitation using the maximum likelihood EM reconstruction algorithm*, IEEE Transactions on Medical Imaging, 13 (1994), pp. 526–537. 84

[16] S. R. Cherry, J. A. Sorensen, and M. E. Phelps, *Physics in Nuclear Medicine*, Saunders, Philadelphia, Pennsylvania, third ed., 2003. 15

[17] G. Chinn and S.-C. Huang, *A general class of preconditioners for statistical iterative reconstruction of emission computed tomography*, IEEE Transactions on Medical Imaging, 16 (1997), pp. 1–10. 55

[18] P. G. Ciarlet, *The Finite Element Method for Elliptic Equations*, North-Holland, Amsterdam, 1978. 24, 25, 26

[19] N. H. Clinthorne, T.-S. Pan, P.-C. Chiao, W. L. Rogers, and J. A. Stamos, *Preconditioning methods for improved convergence rates in iterative reconstructions*, IEEE Transactions on Medical Imaging, 12 (1993), pp. 78–83. 55

[20] R. Courant and D. Hilbert, *Methods of Mathematical Physics*, vol. 1, Interscience Publishers, Inc., New York, New York, 1953. 4, 42

[21] P. J. Davis, *Circulant Matrices*, John Wiley and Sons, New York, 1979. 57

[22] A. R. De Pierro, *A modified expectation-maximization algorithm for penalized likelihood estimation in emission tomography*, IEEE Transactions on Medical Imaging, 14 (1995), pp. 132–137. 4

[23] H. Derin and W. S. Cole, *Segmentation of textured images using Gibbs random fields*, Computer Vision, Graphics and Image Processing, 35 (1986), pp. 72–98. 37

[24] H. Derin and H. Elliott, *Modeling and segmentation of noisy and textured images using Gibbs random fields*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 9 (1987), pp. 39–55. 37

[25] J. Duchon, *Interpolation des fonctions de deux variables suivant le principe de la flexion des plaques minces*, Revue Française d'Automatique, Informatique, Recherche Opérationnelle. Analysis Numérique, 10 (1976), pp. 5–12. 24, 26, 28

[26] ——, *Splines minimizing rotation-invariant semi-norms in Sobolev spaces*, in Constructive Theory of Functions of Several Variables, W. Schempp and K. Zeller, eds., Springer-Verlag, Berlin, 1977, pp. 85–100. 24, 26, 28

[27] H. Erdoğan and J. A. Fessler, *Monotonic algorithms for transmission tomography*, IEEE Transactions on Medical Imaging, 18 (1999), pp. 801–814. 4, 53

[28] J. A. Fessler, *Penalized weighted least-squares image reconstruction for positron emission tomography*, IEEE Transactions on Medical Imaging, 13 (1994), pp. 290–300. 4, 54

[29] J. A. Fessler and S. Booth, *Conjugate-gradient preconditioning methods for shift-variant PET image reconstruction algorithm*, IEEE Transactions on Image Processing, 8 (1999), pp. 688–699. 55

[30] J. A. Fessler, E. P. Ficaro, N. H. Clinthorne, and K. Lange, *Grouped-coordinate ascent algorithms for penalized-likelihood transmission image reconstruction*, IEEE Transactions on Medical Imaging, 16 (1997), pp. 166–175. 4

[31] J. A. Fessler and A. O. Hero, *Space-alternating generalized expectation-maximization algorithm*, IEEE Transactions on Signal Processing, 42 (1994), pp. 2664–2677. 4

[32] ——, *Penalized maximum-likelihood image reconstruction using space-alternating generalized EM algorithms*, IEEE Transactions on Image Processing, 4 (1995), pp. 1417–1429. 4

[33] R. Fletcher and C. M. Reeves, *Function minimization by conjugate gradients*, Computer Journal, 7 (1964), pp. 149–154. 56

[34] L. R. Furenlid, D. W. Wilson, Y. Chen, H. Kim, P. J. Pietraski, M. J. Crawford, and H. H. Barrett, *FastSPECT II: A second-generation high-resolution dynamic SPECT imager*, IEEE Transactions on Nuclear Science, 51 (2004), pp. 631–635. 10

[35] D. Geiger and F. Girosi, *Parallel and deterministic algorithms for MRF's: Surface reconstruction*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 13 (1991), pp. 401–412. 30, 46, 49, 56

[36] D. Geman and G. Reynolds, *Constrained restoration and the recovery of discontinuities*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 14 (1992), pp. 367–383. 5, 43

[37] S. Geman and D. Geman, *Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 6 (1984), pp. 721–741. 5, 30, 31, 35, 38, 49, 62

[38] S. Geman and D. E. McClure, *Statistical methods for tomographic image reconstruction*, Bulletin of the International Statistical Institute, 52 (1987), pp. 5–21. 3

[39] G. Gindi, M. Lee, A. Rangarajan, and I. G. Zubal, *Bayesian reconstruction of functional images using anatomical information as priors*, IEEE Transactions on Medical Imaging, 12 (1993), pp. 670–680. 6, 47, 63, 67, 68, 71, 72, 93

[40] G. Gindi, A. Rangarajan, M. Lee, P. J. Hong, and I. G. Zubal, *Bayesian reconstruction for emission tomography via deterministic annealing*, in Information Processing in Medical Imaging, Springer–Verlag, 1993, pp. 322–338. 6, 47, 56, 62, 63, 67, 68, 71, 72, 93

[41] S. S. Gleason, D. W. Austin, R. S. Beach, R. Nutt, and M. J. Paulus, *A new highly versatile multimodality small animal imaging platform*, in Proceedings of the 2006 IEEE Nuclear Science Symposium and Medical Imaging Conference, 2006, pp. 2447–2449. 16

[42] G. H. Golub and C. V. Loan, *Matrix Computations*, The Johns Hopkins University Press, Baltimore, Maryland, third ed., 1996. 56

[43] P. J. Green, *Bayesian reconstruction from emission tomography data using a modified EM algorithm*, IEEE Transactions on Medical Imaging, 9 (1990), pp. 84–93. 3

[44] J. Gregor, T. Benson, S. S. Gleason, M. J. Paulus, S. D. Figueroa, T. Hoffman, S. J. Kennel, and J. S. Wall, *Approximate volumetric systems models for MicroSPECT*, IEEE Transactions on Nuclear Science, 53 (2006), pp. 2646–2652. 11, 77

[45] W. E. L. Grimson, *From Images to Surfaces*, MIT Press, Cambridge, Massachusetts, 1981. 4, 19, 29, 42

[46] P. C. Hansen, *Analysis of discrete ill-posed problems by means of the L-curve*, SIAM Review, 34 (1992), pp. 561–580. 60

[47] P. C. Hansen and D. P. O'Leary, *The use of the L-curve in the regularization of discrete ill-posed problems*, SIAM Journal of Scientific Computing, 14 (1993), pp. 1487–1503. 60

[48] T. J. Hebert and R. M. Leahy, *A generalized EM algorithm for 3-D Bayesian reconstruction from Poisson data using Gibbs priors*, IEEE Transactions on Medical Imaging, 8 (1989), pp. 194–202. 2, 3, 4, 53

[49] ——, *Statistic-based MAP image reconstruction from Poisson data using Gibbs priors*, IEEE Transactions on Signal Processing, 40 (1992), pp. 2290–2303. 60

[50] F. Heitz and P. Bouthemy, *Multimodal estimation of discontinuous optical flow using Markov random fields*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 15 (1993), pp. 1217–1232. 30

[51] H. M. Hudson and R. S. Larkin, *Accelerated image reconstruction using ordered subsets of projection data*, IEEE Transactions on Medical Imaging, 13 (1994), pp. 601–609. 4, 53

[52] K. Ishizui, T. Mukai, Y. Yonekura, M. Pagani, T. Fujita, Y. Magata, S. Nishizawa, N. Tamaki, H. Shibasaki, and J. Konishi, *Ultra-high resolution SPECT system using four pinhole collimators for small animal studies*, Journal of Nuclear Medicine, 36 (1995), pp. 2282–2286. 10

[53] S. Jan, G. Santin, D. Strul, S. G. Staelens, K. Assié, D. Autret, S. Avner, R. Barbier, M. Bardiès, P. M. Bloomfield, D. Brasse, V. Breton, P. Bruyndonckx, I. Buvat, A. F. Chatziioannou, Y. Choi, Y. H. Chung, C. Comtat, D. Donnarieix, L. Ferrer, S. J. Glick, C. J. Groiselle, D. Guez, P.-F. Honoré, S. Kerhoas-Cavata, A. S. Kirov, V. Kohli, M. Koole, M. Krieguer, D. J. van der Laan, F. Lamare, G. Largeron, C. Lartizien, D. Lazaro, M. C. Maas, L. Maigne, F. Mayet, F. Melot, C. Merheb, E. Pennacchio, J. M. Pérez, U. Pietrzyk, F. R. Rannou, M. Rey, D. R. Schaart, C. R. Schmidtlein, L. Simon, T. Y. Song, J.-M. Vieira, D. Visvikis, R. V. de Walle, E. Wieërs, and C. Morel, *GATE: a simulation toolkit for PET and SPECT*, Physics in Medicine and Biology, 49 (2004), pp. 4543–1034. 11, 80

[54] R. J. Jaszczak, P. H. Murphy, D. Huard, and J. A. Burdine, *Radionuclide emission computed tomography of the head with ^{99m}Tc and a scintillation camera*, Journal of Nuclear Medicine, 18 (1977), pp. 373–380. 10

[55] C. A. Johnson, *Nonlinear Optimization for Volume PET Reconstruction*, PhD thesis, George Mason University, Fairfax, Virginia, 1997. 54

[56] C. A. Johnson, J. Seidel, and A. Sofer, *Interior-point methodology for 3-D PET reconstruction*, IEEE Transactions on Medical Imaging, 19 (2000), pp. 271–285. 4

[57] C. A. Johnson, Y. Yan, R. E. Carson, R. L. Martino, and M. Daube-Witherspoon, *A system for the 3D reconstruction of retracted-septa PET data*

using the EM algorithm, IEEE Transactions on Nuclear Science, 42 (1995), pp. 1223–1227. 4

[58] L. Kaufman, *Implementing and accelerating the EM algorithm for positron emission tomography*, IEEE Transactions on Medical Imaging, 6 (1987), pp. 37–50. 4, 54

[59] ——, *Maximum likelihood, least squares, and penalized least squares for PET*, IEEE Transactions on Medical Imaging, 12 (1993), pp. 200–214. 4, 54

[60] P. E. Kinahan and J. G. Rogers, *Analytic 3D image reconstruction using all detected events*, IEEE Transactions on Nuclear Science, 36 (1989), pp. 964–968. 4

[61] J. Konrad and E. Dubois, *Bayesian estimation of motion vector fields*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 14 (1992), pp. 910–927. 30

[62] K. Lange, *Convergence of EM image reconstruction algorithms with Gibbs smoothing*, IEEE Transactions on Medical Imaging, 9 (1990), pp. 439–446. 3

[63] K. Lange and R. E. Carson, *EM reconstruction algorithms for emission and transmission tomography*, Journal of Computer Assisted Tomography, 8 (1982), pp. 306–316. 4, 51

[64] P. Lascaux and P. Lesaint, *Some nonconforming finite elements for the plate bending problem*, Revue Française d'Automatique, Informatique, Recherche Opérationnelle. Analysis Numérique, 9 (1975), pp. 9–53. 29

[65] R. M. Leahy and X.-H. Yan, *Incorporation of anatomical MR data for improved functional imaging with PET*, in Information Processing in Medical Imaging, G. Goos and J. Hartmanis, eds., 1991, pp. 105–120. 72

[66] Y. G. Leclerc, *Constructing simple stable descriptions for image partitioning*, International Journal of Computer Vision, 3 (1989), pp. 73–102. 46

[67] M. Lee, A. Rangarajan, I. G. Zubal, and G. Gindi, *A continuation method for emission tomography*, IEEE Transactions on Nuclear Science, 40 (1993), pp. 2049–2058. 4, 6, 42, 47, 56, 63, 66, 68, 71, 72, 93, 111, 114

[68] S.-J. Lee, *Bayesian Image Reconstruction in Emission Computed Tomography using Mechanical Models as Priors*, PhD thesis, State University of New York at Stony Brook, Stony Brook, New York, 1995. 65, 66, 69, 70, 71, 72, 74, 113

[69] S.-J. Lee, Y. Choi, and G. Gindi, *Validation of new Gibbs priors for Bayesian tomographic reconstruction using numerical studies and physically acquired data*, IEEE Transactions on Nuclear Science, 46 (1999), pp. 2154–2161. 93

[70] S.-J. Lee, A. Rangarajan, and G. Gindi, *Bayesian image reconstruction in SPECT using higher order mechanical models as priors*, IEEE Transactions on Medical Imaging, 14 (1995), pp. 689–680. 4, 48, 56, 63, 66, 67, 68, 69, 71, 72, 74, 80, 84, 93, 113

[71] S. Z. Li, *Markov random field modeling in image analysis*, Springer-Verlag, San Francisco, third ed., 2009. 30, 33, 34, 35, 36

[72] M. Liévin and F. Luthon, *Nonlinear color space and spatiotemporal MRF for hierarchical segmentation of face features in video*, IEEE Transactions on Image Processing, 13 (2004), pp. 1–9. 37

[73] W. Lin and Y. Liu, *Using a deformable surface model to obtain a shape representation of the cortex*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 29 (2007), pp. 777–792. 37

[74] J. Llacer and E. Veklerov, *Feasible images and practical stopping rules for iterative algorithms in emission tomography*, IEEE Transactions on Medical Imaging, 8 (1989), pp. 186–193. 80

[75] D. Luenberger, *Linear and Nonlinear Programming*, Addison-Wesley, Reading, Massachusetts, second ed., 1984. 54, 56, 68

[76] L. R. MacDonald, B. E. Patt, J. S. Iwanczyk, B. M. W. Tsui, Y. Wang, E. C. Frey, D. E. Wessell, P. D. Acton, and H. F. Kung, *Pinhole SPECT of mice using the LumaGEM gamma camera*, IEEE Transactions on Nuclear Science, 48 (2001), pp. 830–836. 10

[77] T. McInerney and D. Terzopoulos, *Deformable models in medical image analysis: a survey*, Medical Image Analysis, 1 (1996), pp. 91–108. 19

[78] J. Meinguet, *Multivariate interpolation at arbitrary points made simple*, Journal of Applied Mathematics and Physics (ZAMP), 30 (1979), pp. 292–304. 24, 28

[79] J. J. Moré and G. Toraldo, *Algorithms for bound constrained quadratic programming problems*, Numerische Mathematik, 55 (1989), pp. 377–400. 55

[80] ——, *On the solution of large quadratic programming problems with bound constraints*, SIAM Journal on Optimization, 1 (1991), pp. 93–113. 55

[81] E. Mumcuoğlu, R. M. Leahy, and S. R. Cherry, *Bayesian reconstruction of PET images: methodology and performance analysis*, Physics in Medicine and Biology, 41 (1996), pp. 1777–1807. 4, 54

[82] E. Mumcuoğlu, R. M. Leahy, S. R. Cherry, and Z. Zhou, *Fast gradient-based methods for Bayesian reconstruction of transmission and emission PET images*, IEEE Transactions on Medical Imaging, 13 (1994), pp. 687–701. 4, 53, 54

[83] F. Natterer, *The Mathematics of Computerized Tomography*, John Wiley Sons, Chichester, Great Britain, 1986. 12

[84] J. Nocedal and S. J. Wright, *Numerical Optimization*, Springer, New York, New York, second ed., 2006. 56

[85] J. M. Ortega and W. C. Rheinbolt, *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York, 1970. 63

[86] R. B. Paranjape, *Fundamental enhancement techniques*, in Handbook of Medical Imaging, I. N. Bankman, ed., Academic Press, San Diego, California, 2000, pp. 3–18. 77

[87] K. Rohr, *Landmark-based Image Analysis*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 2001. 19

[88] N. U. Schramm, A. K. Wirrwar, F. Sonnenberg, and H. Halling, *Compact high resolution detector for small animal SPECT*, IEEE Transactions on Nuclear Science, 47 (2001), pp. 1163–1167. 10

[89] L. A. Shepp and Y. Vardi, *Maximum likelihood reconstruction for emission tomography*, IEEE Transactions on Medical Imaging, 1 (1982), pp. 113–122. 2, 4, 51

[90] J. Sun, N.-N. Zheng, and H.-Y. Shum, *Stereo matching using belief propagation*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 25 (2003), pp. 787–800. 30

[91] R. Szeliski, R. Zabih, D. Scharstein, O. Veksler, V. Kolmogorov, A. Agarwala, M. Tappen, and C. Rother, *A comparative study of energy minimization methods for markov random fields with smoothness-based priors*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 30 (2008), pp. 1068–1080. 94

[92] D. Terzopoulos, *Multilevel computation processes for visual surface reconstruction*, Computer Vision, Graphics and Image Processing, 24 (1983), pp. 52–96. 4, 19, 20, 22, 29, 42, 93

[93] ——, *Regularization of inverse visual problems involving discontinuities*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 8 (1986), pp. 413–424. 4, 19, 24, 29, 42

[94] A. M. Thompson, J. C. Brown, J. W. Kay, and M. Titterington, *A study of methods of choosing the smoothing parameter in image restoration by regularization*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 13 (1991), pp. 326–339. 60

[95] P. J. M. van Laarhoven and E. H. L. Aarts, *Simulated Annealing: Theory and Applications*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1987. 62

[96] G. Wahba, *Spline Models for Observational Data*, SIAM Press, Philadelphia, 1990. 26, 60

[97] J. S. Wall, S. J. Kennel, M. J. Paulus, S. S. Gleason, J. Gregor, J. Baba, M. Schell, T. Richey, B. O'Nuallain, R. Donnell, P. N. Hawkins, D. T. Weiss, and A. Solomon, *Quantitative high-resolution microradiographic imaging of amyloid deposits in a novel murine model of AA-amyloidosis*, Amyloid: The Journal of Protein Folding Disorders, 12 (2005), pp. 149–156. 9

[98] D. A. Weber, M. Ivanovic, D. Franceschi, S.-E. Strand, K. Erlandsson, M. Franceschi, H. L. Atkins, J. A. Coderre, H. Susskind, T. M. Button, and K. Ljunggren, *Pinhole SPECT: An approach to* in vivo *high resolution SPECT imaging in small laboratory animals*, Journal of Nuclear Medicine, 35 (1994), pp. 342–348. 10

[99] O. Woodford, P. Torr, I. Reid, and A. Fitzgibbon, *Global stereo reconstruction under second-order smoothness priors*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 31 (2009), pp. 2115–2126. 94

[100] A. Yuille, *Energy functions for early vision and analog networks.*, Biological Cybernetics, 61 (1987), pp. 115–123. 45, 46, 49

[101] Z. Zhou, R. M. Leahy, and J. Qi, *Approximate maximum likelihood hyperparameter estimation for Gibbs priors*, IEEE Transactions on Image Processing, 6 (1997), pp. 844–861. 60

Appendices

Appendix A

Derivation of Weak Plate

Following the approach of (67), the M-step objective function takes the following form with the prior energy of the quadratic variation:

$$\begin{aligned} M(f|\hat{f}^k;T) = &\sum_{t\theta}\sum_{ij}\left[(A_{t\theta,ij}f_{ij} - d_{t\theta})\frac{A_{t\theta;ij}\hat{f}^k_{ij}\log\left(f_{ij}\right)}{\sum_{kl}A_{t\theta;kl}\hat{f}^k_{kl}}\right] \\ &+\sum_{ij}\left[-\frac{1}{T}\log\left(\exp\left(-T\beta U_{ij} + \exp\left(-T\beta\alpha\right)\right)\right)\right] \end{aligned} \tag{A.1}$$

where U_{ij} is defined in Equation 4.53. A change of variables in the objective function can be performed by applying the following transformation to each voxel indexed by (i,j):

$$\psi(\sqrt{U_{ij}}, s^h) \stackrel{\text{def}}{=} \xi(s) + \mu(U_{ij} - s). \tag{A.2}$$

where

$$\xi(s) \stackrel{\text{def}}{=} -\frac{1}{T}\log\left(\exp\left(-T\beta s\right) + \exp(-T\beta\alpha)\right) \tag{A.3}$$

and μ is a Lagrangian parameter whose purpose is to force s towards U_{ij} by the action of the Langrange constant term in Equation A.2. As in (67), the definition of $\xi(s)$ depends upon the introduction of a new variable s. Differentiating $\xi(s)$ with respect

to s and setting the result to zero yields

$$\mu = \xi'(s) \tag{A.4}$$

The transformation $\psi(\sqrt{U_{ij}}, s)$, after Lagrange parameter elimination, becomes

$$\psi(\sqrt{U_{ij}}, s) = \xi(s) + \xi'(s)(U_{ij} - s). \tag{A.5}$$

In Equation A.2, the derivation with respect to s is set to zero, which yields

$$\begin{aligned} \frac{\partial^2 \psi(\sqrt{U_{ij}}, s)}{\partial s^2} = 0 \Longrightarrow \quad & \xi'(s) + \xi''(s)(U_{ij} - s) - \xi'(s) = 0 \\ \Longrightarrow \quad & \xi''(s)(U_{ij} - s) = 0 \end{aligned} \tag{A.6}$$

The positivity condition of the second derivative is met at the fixed point $s = U_{ij}$ since $\xi''(s) < 0$, thereby ensuring that the fixed point is a minimum. The second derivative is now

$$\xi''(s) = -T\beta^2 \frac{\exp(-T\beta(s-\alpha))}{\left[1 + \exp(-T\beta(s-\alpha))\right]^2} \leq 0 \tag{A.7}$$

The second derivative condition $\xi''(s) = 0$ occurs only when $\beta \to \infty$ or $s \to \infty$. In practice, the former does not occur and $s = [0, \infty)$; therefore, the second derivative condition is verified and the objective function has two sets of variables f_{ij} and s_{ij} along with the fixed point condition $\hat{s} = U_{ij}$ and $s = U_{ij}$.

Equation A.2 can now be rewritten as

$$\psi(\sqrt{U_{ij}}, s) = U_{ij}\xi'(s) + (\xi(s) - s\xi'(s)). \tag{A.8}$$

Using the transformation

$$z = z(s) = 1 - \xi'(s)/\beta \tag{A.9}$$

gives

$$\psi(\sqrt{U_{ij}}, z_{ij}) = \beta U_{ij}(1 - z_{ij}) + \beta\alpha z_{ij} + \frac{1}{T}(z_{ij}\log z_{ij} + (1 - z_{ij})\log(1 - z_{ij})). \quad \text{(A.10)}$$

When Equation A.10 is minimized with respect to z, the result is

$$z = \frac{1}{\exp(-T\beta(U_{ij} - \alpha))} \quad \text{(A.11)}$$

which corresponds to the expected value of the line process for the weak plate (68; 70). Using Equation A.10, the objective function becomes

$$\begin{aligned} M(f, z \mid \hat{f}, T) = &\sum_{ij}\sum_{t,\theta}\left[(A_{t\theta,ij}f_{ij} - d_{t\theta})\frac{A_{t\theta;ij}\hat{f}^n_{ij}}{\sum_{kl} A_{t\theta;kl}\hat{f}^n_{kl}}\log(f_{ij})\right] \\ &+ \sum_{ij}\left[\beta U_{ij}(1 - z_{ij}) + \alpha z_{ij}\right] \\ &+ \frac{1}{T}\sum_{ij}\left[z_{ij}\log z_{ij} + (1 - z_{ij})\log(1 - z_{ij})\right]. \end{aligned} \quad \text{(A.12)}$$

This is the objective function for the weak plate (68; 70).

The second partial derivatives are calculated to verify that the solution obtained by differentiating with respect to f_{ij} or z_{ij} and setting it to zero is a minimum, it suffices to demonstrate that M is convex with respect to f_{ij} and z_{ij} (68; 70):

$$\begin{aligned} \frac{\partial^2 M(f, z|\hat{f}^n; T)}{f^2_{ij}} &= \frac{1}{f^2_{ij}}\left(\sum_{t\theta} d_{t\theta}\frac{A_{t\theta;ij}\hat{f}^n_{ij}\log(f_{ij})}{\sum_{kl}\mathcal{H}_{t\theta;kl}\hat{f}^n_{kl}}\right) \\ &\quad + 2\beta\left\{(10(1 - z_{ij}) + (1 - z_{i,j+1}) + 3(1 - z_{i,j-1}) + (1 - z_{i+1,j})\right. \\ &\quad \left. +3(1 - z_{i-1,j}) + 2(1 - z_{i-1,j-1})\right\} \geq 0 \\ \frac{\partial^2 M(f, z|\hat{f}^n; T)}{\partial z^2_{ij}} &= \frac{1}{T z_{ij}(1 - z)} \geq 0 \end{aligned} \quad \text{(A.13)}$$

Appendix B

Derivation of Thin-Plate Spline

As in Appendix A, the derivation of the objective function for the thin-plate spline follows the approach of (67). As before, the M-step objective function takes the following form with the prior energy of the quadratic form:

$$\begin{aligned} M(f|\hat{f}^k;T) = \sum_{t\theta}\sum_{ijk}\left[(A_{t\theta,ijk}f_{ijk} - d_{t\theta})\frac{A_{t\theta;ijk}\hat{f}^n_{ijk}\log\left(f_{ijk}\right)}{\sum_{klm}A_{t\theta;klm}\hat{f}^n_{klm}}\right] \\ + \sum_{ijk}\left[-\frac{1}{T}\log\left(\exp\left(-T\beta U_{ijk} + \exp\left(-T\beta\alpha\right)\right)\right)\right] \end{aligned} \tag{B.1}$$

A change of variables in the objective function is performed by applying the following transformation to each indexed by (i,j,k):

$$\psi(\sqrt{U_{ijk}}, s) \stackrel{\text{def}}{=} \xi(s) + \mu(U_{ijk} - s). \tag{B.2}$$

where

$$\xi(s) \stackrel{\text{def}}{=} -\frac{1}{\beta}\log\left(\exp\left(-\beta\lambda s\right) + \exp(-\beta\lambda\alpha)\right) \tag{B.3}$$

where μ is a Lagrangian parameter whose purpose is to force s towards U_{ijk} by the action of the Lagrange constant term. The definition of $\xi(s)$ depends upon the introduction of a new variable s. Differentiating $\xi(s)$ with respect to s and setting

the result to zero yields

$$\mu = \xi'(s) \tag{B.4}$$

The transformation $\psi(\sqrt{U_{ijk}}, s)$, after Lagrange parameter elimination, becomes

$$\psi(\sqrt{U_{ijk}}, s) = \xi(s) + \xi'(s)(U_{ijk} - s). \tag{B.5}$$

In Equation B.5, the derivation with respect to s is set to zero, which yields

$$\begin{aligned} \frac{\partial^2 \psi(\sqrt{U_{ijk}}, s)}{\partial s^2} = 0 \Longrightarrow \quad & \xi(s) + \xi'' U_{ijk} - s) - \xi'(s) = 0 \\ \Longrightarrow \quad & \xi''(s)(U_{ijk} - s) \\ = \quad & 0 \end{aligned} \tag{B.6}$$

The fixed point is at $s = U_{ijk}$ provided that the second derivative condition $\xi''(s) < 0$ is met, thereby ensuring that the fixed point is a minimum. The second derivative is now

$$\xi''(s) = T\beta^2 \frac{\exp(-T\beta(s-\alpha))}{\left[1 - \exp(-T\beta(s-\alpha))\right]^2} \tag{B.7}$$

The second derivative condition $\xi''(s) = 0$ occurs only when $s \to \infty$. As before, the second derivative condition is verified and the objective function has two sets of variables f_{ijk} and s_{ijk} along with the fixed point condition $\hat{s}^h = U_{ijk}$.

Equation B.5 can now be rewritten as

$$\psi(\sqrt{U_{ijk}}, s) = U_{ijk}\xi'(s) + (\xi(s) - s\xi'(s)). \tag{B.8}$$

Using the transformation

$$z = z(s) = 1 - \xi'(s)/\beta \tag{B.9}$$

gives

$$\psi(\sqrt{U_{ijk}}, z_{ijk}) = \beta U_{ijk}(1 - z_{ijk}) + \beta\alpha z_{ijk} + \frac{1}{T}(z_{ijk}\log z_{ijk} + (1 - z_{ijk})\log(1 - z_{ijk})). \tag{B.10}$$

When Equation B.10 is minimized with respect to z, the result is

$$z = \frac{1}{\exp(-T\beta(U_{ijk} - \alpha))} \tag{B.11}$$

which corresponds to the expected value of the line process in Equation 5.48. Using Equation B.10, the objective function becomes

$$\begin{aligned} M(f, z \mid \hat{f}, T) = &\sum_{i,j,k}\sum_{t,\theta}\left[(A_{t\theta,ijk}f_{ijk} - d_{t\theta})\frac{A_{t\theta;ijk}\hat{f}^n_{ijk}}{\sum_{klm} A_{t\theta;klm}\hat{f}^n_{klm}}\log\left(f_{ijk}\right)\right] \\ &+ \sum_{i,j,k}\left[\beta U_{i,j,k}(1 - z_{i,j,k}) + \alpha z_{i,j,k}\right] \\ &+ \frac{1}{T}\sum_{i,j,k}\left[z_{i,j,k}\log(z_{i,j,k}) + (1 - z_{i,j,k})\log(1 - z_{i,j,k})\right]. \end{aligned} \tag{B.12}$$

This objective function is defined in Equation 5.50.

The second partial derivatives are calculated to verify that the solution obtained by differentiating with respect to f_{ijk} or z_{ijk} and setting it to zero is a minimum, it suffices to demonstrate that M is convex with respect to f_{ijk} and z_{ijk}.

$$\begin{aligned} \frac{\partial^2 M(f, z|\hat{f}^n; T)}{f^2_{ijk}} = &\frac{1}{f^2_{ijk}}\left(\sum_{t\theta} d_{t\theta}\frac{A_{t\theta;ijk}\hat{f}^n_{ijk}\log\left(f_{ijk}\right)}{\sum_{lmn}\mathcal{H}_{t\theta;lmn}\hat{f}^n_{lmn}}\right) \\ &+ 2\beta\left\{21(1 - z_{i,j,k}) + 5(1 - z_{i-1,j,k}) + 5(1 - z_{i,j-1,k})\right. \\ &+ 5(1 - z_{i,j,k-1}) + (1 - z_{i+1,j,k}) + (1 - z_{i,j+1,k}) \\ &+ (1 - z_{i,j,k+1}) + (1 - z_{i-1,j-1,k}) + (1 - z_{i-1,j,k-1}) \\ &\left.+ (1 - z_{i,j-1,k-1})\right\} \geq 0 \\ \frac{\partial^2 M(f, z|\hat{f}^n; T)}{\partial z^2_{ijk}} = &\frac{1}{T z_{ijk}(1 - z_{ijk})} \geq 0 \end{aligned} \tag{B.13}$$

Vita

Lloyd Fredrick Arrowood was born on October 2, 1959 in Chattanooga, Tennessee. He attended elementary school in Chattanooga, but moved to Johnson City, Tennessee where he completed his secondary education, graduating from Science Hill High School. He graduated from the University of Tennessee with a Bachelor of Arts degree in Computer Science in 1980 and a Master of Science degree in Computer Science in 1989. He has worked at several United States Department of Energy facilities, including the Oak Ridge National Laboratory, and works currently at the Y-12 National Security Complex in Oak Ridge, Tennessee.