Bayesian Reconstruction In Emission Tomography Using Gibbs Priors

A Dissertation
Presented to the Faculty of the Graduate School
of
Yale University
in Candidacy for the Degree of
Doctor of Philosophy

by
Mindy MinHae Lee

May 1994

ABSTRACT

Bayesian Reconstruction In Emission Tomography Using Gibbs Priors

Mindy MinHae Lee

Yale University

1994

Two methods of improving reconstructions of emission tomography (SPECT and PET) are studied. First, using a Bayesian model we present a mechanism to incorporate prior information derived from a registered anatomical CT/MR image to aid the reconstruction of functional SPECT/PET images. Our method encourages the reconstructed image to be piecewise smooth; the contribution here is the inclusion of a coupling term that influences the creation of edges in the vicinity of significant anatomical edges. The coupling term is modeled in a Gibbs prior distribution. Simulations on mathematical phantoms are presented. Second, we have employed a *continuation method* to develop a new framework for emission tomography reconstructions. A Gibbs prior of piecewise smoothness is used in our Bayesian model. Our Gibbs prior potential function, ϕ^*, consistent with the use of line processes, leads to difficult minimization problems. We present a method wherein the "correct" ϕ^* function is approached through a *sequence* of smooth ϕ functions. The minimum of the objective function corresponding to one member of the ϕ function sequence is used as an initial condition for the minimization of the next, less approximated, stage. A deterministic annealing algorithm with closed–form update equations is derived to implement our continuation method. Simulation results on a mathematical phantom are presented.

Acknowledgment

I would like to thank my two advisors, Gene Gindi and George Zubal, for their continuous guidance and encouragement during the course of this work. I am also grateful to James Duncan for serving on my committee and providing valuable suggestions and support. I also wish to thank Arye Nehorai for helpful comments, and Richard McGowan for taking the time to be my reader. I would like to especially thank Anand Rangarajan. Without his help and enlightening discussion, this thesis would not have been the same. Many thanks to Caroline Meloling for providing administrative assistance, and Mary Lally for much help on impossible-to-follow graduate school rules. Thanks also to Chuck Harrell for technical assistance.

I must also thank current and past fellow students and postdocs (a couple of them newly professored) in Image Processing and Analysis Group: Amir Amini, Amit Chakraborty, François Meyer, Glynn Robinson, Peng-Cheng Shi, Hemant Tagare, Joachim Utans, and Lei Zhang. I am particularly grateful to Larry Staib who patiently answered all my questions except the meaning of life, Volker Tresp for his help on my course work in earlier days, John McEachen, if not for his friendship, then at least for the bass guitar, and Rupert Curwen for having answers for all my LaTeX questions.

I would like to thank all my friends (some still around and some who have moved on to bigger cities and/or to bigger lives) who have provided priceless friendships and made my life in New Haven fun and exciting. Special thanks to Craig Copeland, Meredith Goldsmith, Leigh Coral Harris, David Houston, David Kinsella, Jim Monckton, and Lara Stables. Most of all, I can not thank enough Chris Alliegro, who spent many weekends proof-reading this thesis (but don't blame him for any errors), for his continuous emotional support and many free meals.

Finally, I would like to thank my brother, sisters and brother-in-law for their support and encouragement. I am eternally indebted to my parents for their intellectual influence and emotional (not to mention financial) support. I dedicate this thesis to my parents.

This work was supported in part by the US Department of Energy grant DE FG0288ER60724.

"My insignia is a guy surfing on an enormous wave of lava... I have it tattooed on my heart. And I don't mean on the skin of my chest over my heart. I mean tattooed on the organ itself. It's illegal in the States–I had to go to Mexico. It's called visceral tattooing. They have to open you up. They use an ink that contains a radioactive isotope so that the tattoo shows up on X-rays and CAT scans."

Mark Leyner, *"Et Tu, Babe"*

Contents

List of Tables

List of Figures

Chapter 1

Introduction

In nuclear medicine, a pharmaceutical tagged with a radioactive isotope (a radiopharmaceutical) is injected into the human body. By choosing a pharmaceutical which has preferential affinity to a certain type of tissue, a radiopharmaceutical can be delivered to an organ of interest. The spatial distribution of the pharmaceutical in the body can then be mapped by detecting the gamma rays emitted in the radioactive decay of the isotope.

Unlike x-ray techniques, where the radiation source is external to the patient, the source of radiation in nuclear medicine is located within the body. This is referred to as emission mode (an imaging system with an external source of radiation is referred to as transmission mode). The spatial distribution of the radioactive pharmaceutical within a given organ typically yields information about the physiological function of the organ, whereas an x-ray image typically yields anatomical information. Thus nuclear medical imagery, mapping the distribution of emitted radioactivity, is used to localize the malfunctioning (diseased) areas of an organ by measuring abnormal concentrations of radioactivity. A chest x-ray, for example, provides an anatomical map of the thorax, whereas using a radioactive gas such as ^{133}Xe in an emission mode procedure enables study of the gas ventilation of the patient's lungs: the patient inhales the gas and the radioactive distribution within the lungs is mapped by detecting the gamma rays emitted.

A conventional nuclear medicine image is a two-dimensional collapsed view of a three-dimensional object similar to an x-ray radiograph. It is often impossible, however, to identify certain features with only a few conventional two-dimensional views because all depth information is superimposed onto a single plane. Using a technique called tomography we can overcome the superposition of information and obtain cross-sectional views of a patient. In nuclear medicine, the detection of radioactive emission and the subsequent cross sectional image reconstruction procedures are collectively known as emission tomography. This thesis deals primarily with reconstruction techniques developed for use in Single Photon Emission Computed Tomography (SPECT) and Positron Emission Tomography (PET), which are discussed in more detail below.

1.1 Tomography in Medicine

Tomography in medicine is a method used to view or image a selected cross section of a patient without surgical intervention, and is a process to define the three-dimensional distribution of information of interest in the body.

Due to a lack of three-dimensional displaying techniques, the images are usually presented as a series of two-dimensional slices. Before the introduction of computers in tomography, classical tomographic technique imaged a cross section of a patient by blurring other planes. This was done by selecting a particular depth of interest by mechanical focusing of the detector system and, thus blurring other depth planes. The revolution of modern computers, however, has enabled us to obtain tomographic images of narrow cross sections of the body without interference (blurring) from neighboring layers. This technique is known as computed tomography. In computed tomography, a set of projection data taken around the body of a patient is used to image cross sections. Below, we briefly review widely used modern diagnostic tomographic imaging methods in medicine.

The first medical scanner to perform x-ray computed tomography (x-ray CT - commonly known as CAT scan) was developed by Hounsfield in the early 1970s. In x-ray CT, a radiation source and its detector unit rotate together around a patient's body [2]. At each rotation angle, the source irradiates the patient with a parallel (or fan) beam of x-rays. The rays which traverse the body without attenuating are collimated and collected by the detection system. Lower particle counts in the detector indicate regions of high attenuation. The plane from the source to the detector determines the cross section of the patient's body. The set of detector measurements for a given plane is then used to reconstruct this cross section. The detector measurements are often referred to as projections since they represent a projection of attenuation information at each angle. The high signal to noise ratio in CT together with well engineered collimation and mechanics allows high resolution (0.5 - 1.5 mm) images. The images correspond to regions of differing attenuation coefficients of the constituent tissue.

Kuhl [23] first introduced tomographic approaches in rotational scanning of the liver in nuclear medicine in 1958. Over the years, tomographic imaging in nuclear medicine has evolved into two distinct categories: SPECT and PET [11]. Since SPECT and PET both image via emission mode, they are collectively referred to as emission tomography, whereas x-ray CT is referred to as transmission tomography.

In emission tomography, a pharmaceutical tagged with a radioactive isotope is either injected, swallowed, or inhaled and delivered to a target organ in the body. SPECT uses gamma-emitting radioisotopes, such as ^{99m}Tc (the meta-stable isotope of technetium), that emit a single photon [9]. Emitted single photons are most commonly detected by a system with parallel (or fan-beam) lead collimators. For parallel collimators, many sets of parallel line integrals of the source distribution, referred as parallel projections, are obtained by rotating the detection system around the patient. A reconstruction algorithm is then applied to the projection data to estimate the three-dimensional activity distribution. In PET, a positron-emitting radioisotope, such as ^{68}Ga, is used. An emitted positron annihilates with a nearby electron creating two 511 KeV photons traveling in (nearly) opposite directions. Using an annihilation coincidence detection scheme [41], pairs of photons detected simultaneously are counted. When a coincidence event is detected, the annihilation event has occurred somewhere along the line connecting the two points of detection. Since the localization in PET is obtained electronically (coincidence detection) instead of using physical collimators, the detection efficiency in PET is high compared to the one in SPECT (roughly, one million counts per second for PET and ten thousand counts per second for SPECT). Another advantage of PET is labelling naturally occurring elements (^{11}C,^{13}N,^{15}O, or ^{18}F) in biomolecules. PET, however, requires isotopes with half-lives on the order of tens of minutes; this necessitates an expensive cyclotron for on-site source production. On the contrary, SPECT sources are relatively long lived. For instance, ^{99m}Tc, a frequently used radioisotope in SPECT, has a half life of six hours and can thus be purchased from off-site vendors. Consequently, only a handful of large research institutes are equipped with PET scanners whereas most major hospitals and medical centers are equipped with SPECT scanners. The typical spatial resolutions of SPECT is 7 - 8 mm and of PET 5 - 6 mm. Simple SPECT and PET detector systems are depicted in Figure 1.1.

Although its data collection and reconstruction methods are very different from the projection method described above, another widely used tomographic technique is magnetic resonance imaging (MRI), also known as nuclear magnetic resonance (NMR) [39]. It was proposed in 1973 by Lauterbur [28] and by Mansfield and Grannell [37]. Unlike x-rays that interact with atomic electrons, low frequency radio waves are used to interact with the spin of atomic nuclei in the human body. A patient is placed in a static magnetic field that is then subject to pulsed electromagnetic energy in the radio-frequency (RF) range. This results in measurable changes in the net magnetic moment of certain atomic nuclei. Measurements of spin density and other tissue related parameters such as spin-lattice and spin-spin relaxation times yields information on tissue type. MR images based on discrim-

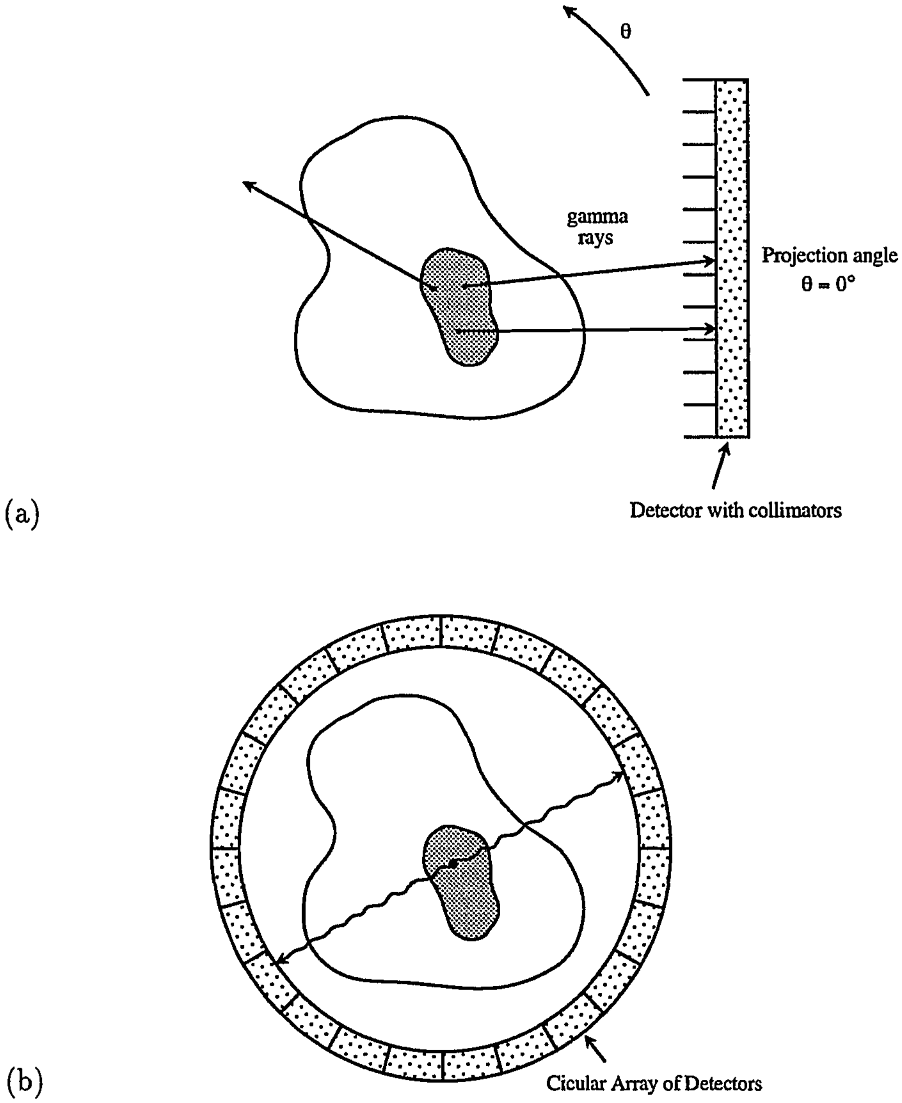


Figure 1.1: Simple Detection Schemes. (a) SPECT. (b) PET.

inating hydrogen proton densities of tissue are commonly used, where the signal results from mainly from the hydrogen in water. Since soft tissues are made up mainly of water, the level of water content can be measured in hydrogen proton MR imaging. These MR images typically provide high resolution (0.5 - 1.5 mm) anatomical information.

Emission tomography is often used in clinical diagnosis either to detect specific disease areas or to study particular physiological functions of an organ. Its ability to obtain physiological and biochemical information is important since physiological and biochemical changes often precede anatomical changes due to pathological conditions. Kuhl *et al.* [24], for example, used emission tomography in neurology to study both regional blood perfusion and metabolism in the brains of patients with partial epilepsy. For the majority of the patients, x-ray CT scans appeared normal. Their PET studies, however, showed increased blood flow and metabolism within epileptic focus during seizure activity. In psychiatry, emission tomography may also provide information on underlying biochemical/physiological defects in certain mental illnesses [1].

1.2 Thesis Outline

In Chapter 2, we present a brief overview of the physics, the instruments, and the imaging characteristics involved in SPECT and PET imaging. Once we obtain the detector measurements (projection data) of SPECT or PET, our task is to estimate, or reconstruct, the three dimensional source distribution. In Chapter 3, we discuss some reconstruction methods used in emission tomography. We present two frameworks: deterministic and probabilistic image reconstruction techniques. Often, various diagnostic imaging methods are used together to make an accurate diagnosis of an illness. In these cases, both anatomical and functional images of the organ of interest are available. Due to the low signal-to-noise ratio in emission tomography compared to that in CT or MR, emission tomography generally produces lower resolution reconstructed images than those of CT/MR. We believe that the high resolution anatomical information obtained in CT/MR scans can aid the reconstruction of functional images of emission tomography if we have a proper way of incorporating the information in our reconstruction problem. In Chapter 4, we investigate a means of fusing the anatomical information obtained from corresponding(registered) CT/MR images with with functional images of emission tomography using a Bayesian probabilistic approach. In Chapter 5, we propose a new deterministic reconstruction framework using a continuation method. We also present the derivation of our deterministic annealing algorithm. Finally, in Chapter 6, we propose possible future directions for SPECT and PET and concluding words.

Chapter 2

Physics and Instruments of SPECT and PET

2.1 Poisson Distribution

In radioactive decay, the total number of gamma rays emitted per unit time by a radioactive source follows a Poisson distribution [49]. The general form of Poisson distribution can be written:

$$\Pr(X = x) = \frac{\mu^x e^{-\mu}}{x!},$$

where $\Pr(X = x)$ is the probability of a random variable, X, of mean, μ, having the value x. The sum of independent Poisson random variables follows a Poisson distribution, with the mean equal to the sum of the means of the individual Poisson distributions. A radioactive source can, thus, be thought of as a three dimensional array of independent Poisson distribution. A detector measurement that counts the gamma emitted by a source follows a Poisson distribution that is the sum of many source distributions. Individual detector measurements in a detector system are independent of all others since each detected photon results from an independent atomic decay and is detected in only one detector bin. Because the mean and the variance of a Poisson distribution are identical, the signal-noise ratio in the measurement data is proportional to the square root of the detector measurement counts. Poisson noise is the dominant noise in the reconstruction images of SPECT and PET.

2.2 Attenuation

To first order, a detector measurement corresponds to the sum of decay activity within pixels along a line selected by a collimator. In practice, however, each decay photon has some probability of interacting, or scattering, in the patient's body before reaching the detection system. Below, we briefly review the physical mechanisms involved in low energy photon interactions. Low energy (approximately 100 KeV) photons (gamma rays and x-rays) interact with matter in three principal ways [22]: Compton scattering, photoelectric effect, and pair production. Compton scattering is a photon-electron interaction in which a photon collides with a free or nearly free electron, loses a part of its energy to the electron and then scatters in a new direction. In the photoelectric effect, a photon is absorbed by an atom, and an electron, called a photoelectron, is emitted from an inner orbit of the atom. Pair production occurs when a photon of energy greater than 1.022 MeV is completely absorbed by the nucleus of an atom and a electron-positron pair is created. The positron (which is the antiparticle of the electron) quickly annihilates with a nearby atomic electron, producing two 511 KeV photons traveling in opposite directions. The source photons in both SPECT and PET are at energy levels considerably less than 1.022 MeV required for pair production. A commonly used isotope in SPECT, ^{99m}Tc, releases gamma rays with an energy of 140 KeV. The photons created from the positron annihilation in PET have energies of 511 KeV. Only Compton scattering and the photoelectric effect are, therefore, significant in SPECT and PET.

Because of these interactions, the number of photons detected differs from the number of photons emitted. The level of attenuation is a function of the type of matter (tissue) traversed. The attenuation coefficient of a given material is measured by the number of photons removed from a collimated (pencil) beam, either through absorption or scattering, during traversal through the material. Since in SPECT and PET only Compton scattering and the photoelectric effect are significant, the total linear attenuation coefficient can be written

$$\mu_{\text{total}} = \mu_{\text{photoelectric}} + \mu_{\text{Compton}}. \tag{2.1}$$

Let N_0 be the original number of photons and $N(x)$ represent the number of photons that travel a distance x in uniform matter without being attenuated. With a constant attenuation coefficient (one type of tissue), μ, $N(x)$ can be written:

$$N(x) = N_0\, e^{-\mu x}. \tag{2.2}$$

Since the body is composed of different types of tissues, there can be different attenuation coefficients involved. Equation 2.2 then becomes:

$$N(x) = N_0\, e^{-\int_0^x \mu(x')dx'}. \tag{2.3}$$

As seen in Equation 2.3, attenuation increases with the depth of the source in the patient. Without attenuation corrections, the reconstructed image will appear less radioactive than the actual distribution. This can cause ambiguity of projection measurements in SPECT, i.e. a strong source at a great distance from the detector and a weak source nearer to the detector can yield the same measurements. The attenuation problem in PET is less serious than in SPECT.

If the attenuation coefficient is a constant, then the attenuation of the annihilation photons depends on the thickness of the body and is independent of position between detectors. This enables reasonable estimation of the attenuation in the cross section being imaged if we know the shape and location of the cross section. The brain, for example, is treated as a region of constant attenuation coefficient. Further, the attenuation coefficients of the human body vary less for annihilation photons than for lower energy photons in SPECT. In water, for example, the attenuation coefficient of a 140 KeV photon is 0.15 cm^{-1} while the attenuation coefficient of a 511 KeV photon is approximately 0.095 cm^{-1}. In a brain of radius 10cm, roughly 20% of 140 KeV photons will be lost to attenuation.

To make accurate attenuation corrections, we must have knowledge of the radioactive distribution at every location within the patient. The distribution of radioactivity is, however, what we are trying to estimate from our detector measurements. A simple approximation, for instance, might be to assume that the patient's contour is an ellipse and that the attenuation of the tissue is uniform. Chang [6] proposed a postcorrection method in which the reconstructed image is folded with an attenuation correction matrix, with the inverse of the average measured attenuation of a source at each location as the correction factor. In an iterative correction method, the corrected image can be reprojected and compared with the measured data by subtraction. From the error projection, we can construct an error matrix and combine it with the corrected image to create a second corrected image, and so on.

2.3 Instruments and Imaging Characteristics of PET and SPECT

The Anger camera is a commonly used instrument to detect emitted photons in SPECT systems [53]. A typical Anger camera consists of a collimator, scintillation crystal and on the order of 30 to 100 photomultiplier tubes. Each photomultiplier also consists of two major elements: a photocathode and an electron multiplier. Collimators are used to select photons at a certain angle at the face of the camera.

For parallel collimators, only the photons arriving perpendicular to the camera are detected.

When a photon hits the surface of the scintillation crystal, it interacts with the atoms of the scintillator and (near) visible light is emitted. When this visible light strikes a photocathode, it releases electrons via the photoelectric effect. The electron signal is amplified by the photomultiplier tube and then fed to the detection electronics to determine the energy of the incident photon and its location in the discretized detector space. The signal-to-noise ratio is very poor from the low photon collection efficiency of the collimators (typical photon counts in a detector bin is roughly 20–30). Photon collection efficiencies for collimators are typically on the order of 10^{-3}.

A ring-shaped (hexagonal, octagonal, circular) array of detectors is often used to count the annihilation gamma rays in PET systems [42]. Frequently used detectors consist of scintillation crystal and a photomultiplier tube. Mechanical collimators are not necessary since two coincident photons determine the line of the positron source location. A higher detection sensitivity is achieved in PET due to the absence of collimators.

In both systems, once a photon is detected and while it is being processed through the electronics, no other photons can be counted in the hit detector crystal. The amount of time needed to process an event in the detector electronics is called deadtime [53]. Even though deadtimes per detector crystal in both PET and SPECT electronics are roughly the same, PET produces a significantly higher integrated counting rate since its array of detectors enable multiple detecting processes.

In SPECT, the spatial resolution is affected by both statistical noise and inaccuracy caused by the imaging system itself. The Anger camera introduces two kinds of blurring: collimator resolution caused by the finite size of collimator holes and intrinsic resolution primarily due to a diffusion of light photons in the crystal. Other inaccuracies may be caused by non-uniformity of camera system response and the misalignment of the center of rotation used in the reconstruction process.

Some of the processes limiting the accuracy of localization of the source particular to PET are: 1) The finite range of positrons before annihilation: a positron travels a short distance before annihilation occurs which makes the location of annihilation slightly different than the location of the positron source, 2) Non-colinear photon emission: the two photons also are not emitted at exactly 180° with respect to each other and 3) background noise in PET: primarily accidental and scatter coincidences. Accidental coincidences occur when photons originating from two independent positron annihilations arrive at the detector pair within the coincidence resolving time. Scatter coincidence occurs when one or both of the photon pair forward-scatter. The line of detection no longer passes through the original annihilation site, causing position resolution errors.

In SPECT, the scattered photons usually have significantly lower energies than the primary photons. Thus the detector system usually uses an energy window to reject photons with less than a certain level of energy to avoid counting scattered photons without significant loss of sensitivity. In PET, however, the photons are at higher energies than those counted in SPECT, and thus lose comparatively less energy by scattering. It is, therefore, much more difficult to cut away the scattered photons without losing a significant number of primary photons. Patient motion during imaging also affects the spatial resolution in both SPECT and PET.

There are also several problems involved with the dynamic change of the pharmaceutical [40] in emission tomography. Often, we assume that the radiopharmaceutical distribution in the body is constant during imaging time. However, the injected pharmaceutical re-distributes within the body over time. To scan the patient, it typically takes about 15 minutes in PET and 30 to 45 minutes in SPECT to collect a significant amount of data. The spatial distribution of the pharmaceutical, however, can change in seconds or minutes. In addition, the radioactive metabolic decaying processes causes the amount of radioactive material to decrease in time. If we could obtain a significant amount of measurements for reconstruction in a shorter imaging time, we could minimize the error due to dynamic change of the radiopharmaceutical within the patient's body.

Chapter 3

Reconstruction Methods

In this section, we describe the techniques used to reconstruct a set of projection data. The result is a map of the radiopharmaceutical distribution in the given cross section of the patient's body. Although many of the reconstruction techniques discussed in this thesis are applicable to both SPECT and PET, the remainder of this thesis deals only with SPECT.

In general, we can group the reconstruction methods into two general approaches: deterministic and probabilistic. In a deterministic approach, an error term is minimized directly whereas in a probabilistic approach, a probability of interest is maximized.

Parallel-hole collimator projection measurements can be represented mathematically in the continuous domain by the following line integral :

$$P_\theta(t) = \int_{Ray} f(x,y)\,ds = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,\delta(x\cos\theta + y\sin\theta - t)\,dx dy \qquad (3.1)$$

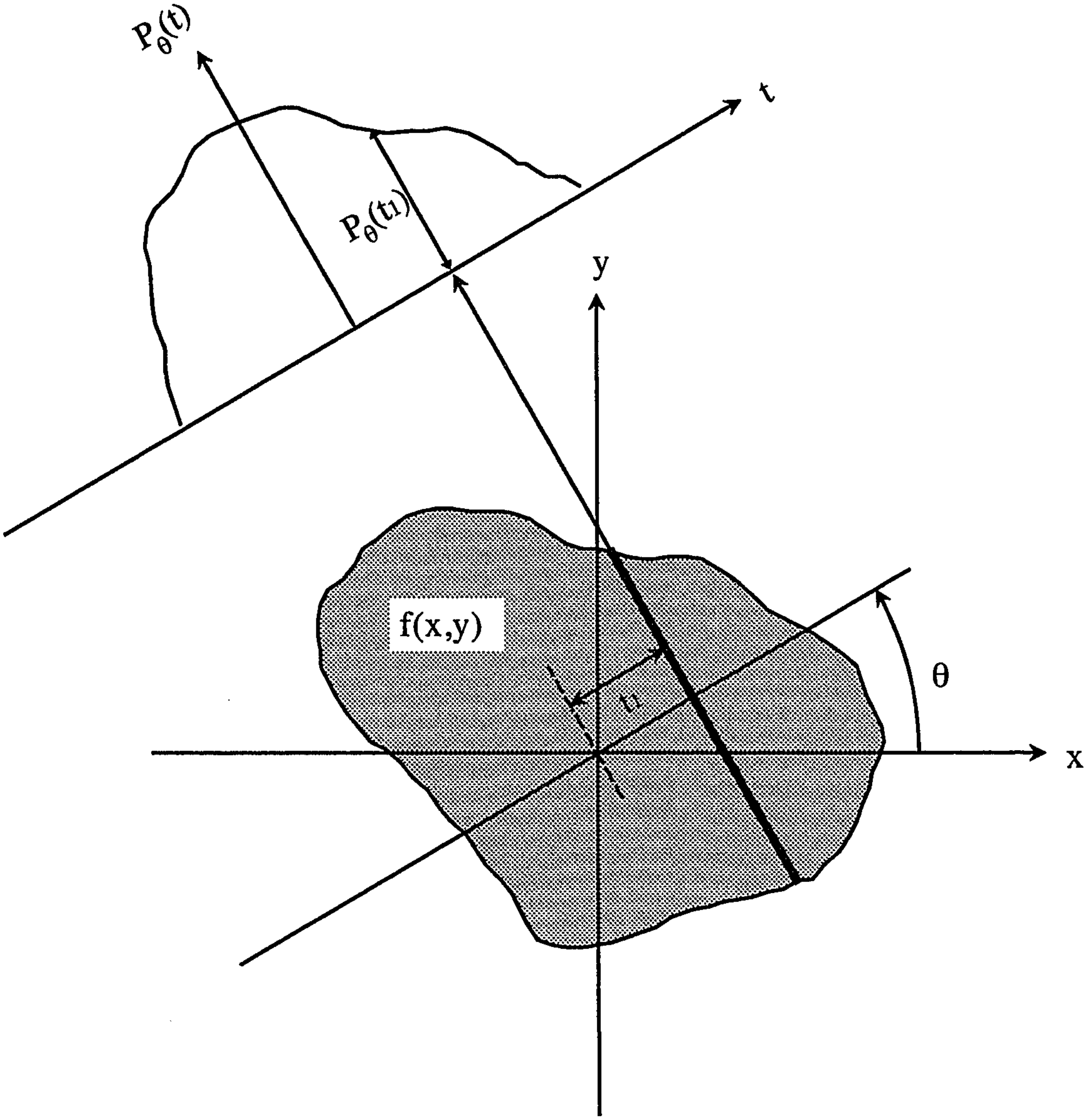


Figure 3.1: Forward Projection Measurement. The upper figure, $P_\theta(t)$, represents the forward projection of the radioactive distribution, $f(x, y)$. $P_\theta(t_1)$ is a particular projection measurement along the dense line in $f(x, y)$ at a distance t_1, from the origin and perpendicular to the line given by θ.

where $f(x,y)$ represents the intensity of emitted radiation at location (x,y), θ is a given projection angle, and t is the detector location. The function $P_\theta(t)$ is the projection of $f(x,y)$ at angle θ, which is the sum of $f(x,y)$ values along a line $t = x\cos\theta + y\sin\theta$, and is called the *Radon Transform* of $f(x,y)$. This is depicted in Figure 3.1.

In emission tomography, however, the radiation measured is not simply the sum of $f(x,y)$ along a ray, but the attenuated sum of $f(x,y)$. If $\mu(x,y)$ represents the attenuation coefficient of a tissue at (x,y), then the *attenuated Radon Transform* of $f(x,y)$ is:

$$P_\theta(t) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\, e^{-\int_{Ray}\mu(x',y')\,dx'dy'}\, \delta(x\cos\theta + y\sin\theta - t)\, dxdy \qquad (3.2)$$

Although the projection process is described in Equation 3.1 and Equation 3.2 in continuous space for mathematical convenience, the actual measured projection data are sampled in discrete space and the source intensity is also estimated in discrete space. The projection data, $\mathbf{g}$, and the source intensity, $\mathbf{f}$, are treated as vector fields, allowing us to use the tools of linear algebra. In Section 3.1, $\mathbf{g}$ and $\mathbf{f}$ denote deterministic quantities for the sake of convenience. For the rest of the thesis, $\mathbf{g}$ and $\mathbf{f}$ will be treated as random quantities.

3.1 Deterministic Framework

In the deterministic framework, the image recovery methods are deterministic. Often, noise is not modeled in the system and the solutions (reconstructed images) follow directly from the projection data.

In forward transformation, a transformation matrix maps the image space to the projection space. Without accounting for noise, it is written

$$g_{t\theta} = \sum_{ij} A_{t\theta;ij} f_{ij} \tag{3.3}$$

where the forward transformation matrix element is $A_{t\theta;ij}$, f_{ij} is the image space, and the projection measurements are represented by $g_{t\theta}$. The reconstruction problem then naturally involves inverting the matrix, A, to obtain the estimate, $\hat{\mathbf{f}}$, of the original source distribution, $\mathbf{f}$, i.e.

$$\hat{\mathbf{f}} = A^{-1}\mathbf{g}.$$

The forward transformation matrices, however, are usually singular or nearly singular: for a given set of projection data, $\mathbf{g}$, there may be *many* (or no) solutions, $\mathbf{f}$, and, therefore, the inverse, A^{-1}, may not exist at all. Even in the case where A^{-1} exists, its computation can become prohibitively expensive. In addition, the reconstruction problem may become ill-posed due to Poisson noise in the projection data.

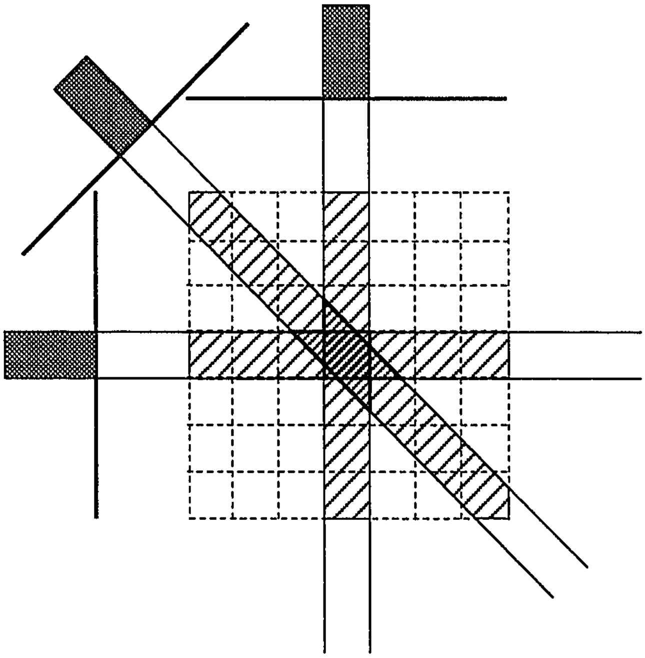


Figure 3.2: Back Projection

The back projection operation can be thought of as the "opposite" process of forward projection. Here, each measured projection value is added to every grid location of the reconstructed source distribution that is viewed by that detector element, as illustrated in Figure 3.2. In general, back-projection of a given forward transformation matrix A can be expressed in summation form as

$$\hat{f}_{B\,ij} = \sum_{t\theta} A_{ij;t\theta} g_{t\theta}. \tag{3.4}$$

In vector notation, the back projection process amounts to applying the transpose of the matrix A to the projection data vector $\mathbf{g}$,

$$\hat{\mathbf{f}}_B = A^T\mathbf{g}. \tag{3.5}$$

Note that $A^T \neq A^{-1}$. The estimate $\hat{\mathbf{f}}_B$ obtained from the back-projection procedure alone has a star-like blurring of the original distribution $\mathbf{f}$. The blurring is caused by the back-projection process and the star pattern is due to the finite number of projection angles used. To correct the blurring, the projection data is filtered with a ramp function in the frequency domain and then the filtered data is back-projected [48]. This is the most commonly used reconstruction method in the clinics and is referred to as filtered back-projection(FBP). Since the ramp filter tends to amplify the high frequency values where signal-to-noise is low, modified ramp filters which roll at high frequency, are usually used.

There are also iterative deterministic approaches [2] that start with a first estimate for $\mathbf{f}$ and correct the error at every step until convergence. An error term at the n^{th} iteration, $\mathbf{e}^n$, can be defined as

$$\mathbf{e}^n = \mathbf{g} - \mathbf{g}^n = \mathbf{g} - A\hat{\mathbf{f}}^n, \tag{3.6}$$

where $\mathbf{g}$ is the measured projection, $\mathbf{g}^n$ is the forward projection of the current estimate, $\hat{\mathbf{f}}^n$, and A is the forward projection operator. As the error vector, $\mathbf{e}$, approaches zero, the estimated solution, $\hat{\mathbf{f}}^n$, approaches the true solution. At every iteration, a new estimate can be calculated by back-projecting the error term and subtracting it from the the current estimate such that

$$\hat{\mathbf{f}}^{n+1} = \hat{\mathbf{f}}^{n} - \sigma A^{T}\mathbf{g} \tag{3.7}$$

where σ is an acceleration parameter with typical value $0 < \sigma < 1$. The iterations are carried out until the error term converges.

In the Algebraic Reconstruction Technique (ART), projection points are addressed one at a time and the image estimate is updated around the projection angle. Because pixel elements are updated at the end of the ray-sum, ART is sometimes called the ray-by-ray reconstruction method. The simultaneous Iterative Reconstruction Technique (SIRT) is another iterative method. In SIRT, the pixel elements are addressed one at a time using data from all the projection points simultaneously and corrected one at a time. SIRT is sometimes called the point-by-point reconstruction method.

3.2 Probabilistic Framework

In the probabilistic framework, image reconstruction involves maximizing a certain probability of interest.

In Equation 3.3, we assume that projection measurements can be made with arbitrary precision. Due, however, to the random nature of radioactive decay, statistical variation or noise is always presents in actual projection data. Since noise is a random variable, each set of projection measurements is different every time, even with the same $\mathbf{f}$. We can not recover $\mathbf{f}$ exactly from $\mathbf{g} = A\mathbf{f}$, regardless of how precise the measurement of $\mathbf{g}$ is.

Table 3.1 summarizes our notational definitions. Lowercase bold quantities denote 2-D vector fields and the corresponding lowercase, italicized quantities denote the elements of the vector field. Similarly, uppercase bold quantities denote 2-D random fields and uppercase italicized quantities denote the corresponding random variables of the random field. For example: $\mathbf{F}$ is a random field, $\mathbf{f}$ is a vector field which is a particular instance of the random field, and f_{ij} is an element of $\mathbf{f}$ corresponding to the location (i, j), and F_{ij} is a random variable corresponding to the location (i, j). The expression $\Pr(\mathbf{F} = \mathbf{f})$ denotes the probability that the random field $\mathbf{F}$ takes the value $\mathbf{f}$.

3.2.1 Maximum Likelihood Estimate

The maximum likelihood (ML) estimate attempts to find an $\hat{\mathbf{f}}$ that maximizes the likelihood function, $\Pr(\mathbf{G} = \mathbf{g}|\mathbf{f})$. In the ML estimate, $\mathbf{f}$, is considered a nonrandom and unknown parameter [55]. For the sake of notational consistency, however, we use the probabilistic notation for $\mathbf{f}$. In general, the likelihood function models noise of the system from which data is acquired. The solution for the ML estimate can be written

$$\hat{\mathbf{f}}_{\mathrm{ML}} = \arg\max_{\mathbf{f}} \Pr(\mathbf{G} = \mathbf{g}|\mathbf{f}).$$

Table 3.1: Symbols used throughout the thesis

$\mathbf{X}$	Random field
X_{ij}	A random variable corresponding to the the location (i,j) of the random field, $\mathbf{X}$
$\mathbf{x}$	A vector field which is a particular instance of the random field, $\mathbf{X}$
x_{ij}	An element of $\mathbf{x}$ corresponding to the location (i,j)
$\mathbf{f}, \mathbf{F}$	2–D source function, associated random field
$\hat{\mathbf{f}}_{\mathrm{MAP}}$	MAP estimate for $\mathbf{f}$
$\mathbf{g}, \mathbf{G}$	Projection data, associated random field
$\mathbf{c}, \mathbf{C}$	Complete data, associated random field
$E_{\mathbf{X}}\{\cdot\}$	Expected value over a random field $\mathbf{X}$
$\Delta^v f_{ij}$	Vertical gradient $f_{i+1,j} - f_{ij}$
$\Delta^h f_{ij}$	Horizontal gradient $f_{i,j+1} - f_{ij}$
$\mathbf{l}, \mathbf{L}$	Line process, associated random field for line process
$\mathbf{l}^h, \mathbf{l}^v$	Horizontal, vertical line processes
$\hat{\mathbf{l}}_{\mathrm{MAP}}$	MAP estimate for $\mathbf{l}$
$\mathcal{H}_{t\theta;ij}$	Forward projection operator
$\bar{g}_{t\theta}$	Expected detector bin count

The likelihood function is the forward projection model prescription for generating the observed projection data, g, from the underlying deterministic source distribution, **f**. The number of photons detected in t at a projection angle, θ, has a Poisson distribution with mean $\sum_{ij} \mathcal{H}_{t\theta;ij} f_{ij}$. The linear forward matrix element, $\mathcal{H}_{t\theta;ij}$, denotes the probability of a photon emitted from the location (i,j) being detected in the detector bin (t,θ). Note that $\mathcal{H}$ is a probability matrix where as A in Section 3.1 is a deterministic matrix. Since each detector bin's counts are independent of all other detector bin counts, as discussed in Section 2.1, the likelihood, $\Pr(\mathbf{G} = \mathbf{g}|\mathbf{f})$, can be expressed as a product of independent Poisson distributions:

$$\Pr(\mathbf{G} = \mathbf{g}|\mathbf{f}) = \prod_{t\theta} \frac{\left(\sum_{ij} \mathcal{H}_{t\theta;ij} f_{ij}\right)^{g_{t\theta}} e^{-\sum_{ij} \mathcal{H}_{t\theta;ij} f_{ij}}}{g_{t\theta}!}. \tag{3.8}$$

In Equation 3.8, $g_{t\theta}$ is the number of counts detected in bin t at projection angle θ, and $\sum_{ij} \mathcal{H}_{t\theta;ij} f_{ij}$ is the expected number of counts in (t,θ) for a particular source distribution, **f**. Note that $\mathcal{H}_{t\theta;ij}$ can account for physical effects such as scattering and attenuation, as well as response and geometrical factors associated with the detector for PET and SPECT. It cannot, however, include nonlinear effects such as detector dead time and random coincidences in PET.

Since the logarithm is a monotonic function, maximizing the likelihood is equivalent to maximizing the log-likelihood. Using Equation 3.8, we have:

$$\log \Pr(\mathbf{G} = \mathbf{g}|\mathbf{f}) = \sum_{t\theta} \left[g_{t\theta} \log(\sum_{ij} \mathcal{H}_{t\theta;ij} f_{ij}) - \sum_{ij} \mathcal{H}_{t\theta;ij} f_{ij} - \log(g_{t\theta}!) \right]. \tag{3.9}$$

The term $\log(g_{t\theta}!)$ can be dropped for the maximization since it is independent of $\mathbf{f}$. The ML estimate, $\hat{\mathbf{f}}_{\mathrm{ML}}$, is then given by

$$\begin{aligned} \hat{\mathbf{f}}_{\mathrm{ML}} &= \arg\max_{\mathbf{f}} \log \Pr(\mathbf{G} = \mathbf{g}|\mathbf{f}) \\ &= \arg\max_{\mathbf{f}} \left[\sum_{t\theta} g_{t\theta} \log(\sum_{ij} \mathcal{H}_{t\theta;ij} f_{ij}) - \sum_{t\theta} \sum_{ij} \mathcal{H}_{t\theta;ij} f_{ij} \right]. \end{aligned} \tag{3.10}$$

3.2.1.1 Expectation Maximization (EM)

One of the approaches to performing the ML estimate is the expectation maximization (EM) technique proposed by Dempster *et al.* [10]. Shepp and Vardi [50], and independently Lange and Carson [27], applied the EM technique to emission tomography. The EM approach uses what is known as an incomplete/complete data formulation. The complete data set $c_{t\theta;ij}$, denotes the number of photons detected at bin (t, θ) that are emitted from pixel (i, j). The incomplete data set $\mathbf{g}$, where $g_{t\theta}$ is the number of photons counted in bin (t, θ), is what we actually observe in emission tomography. The complete data set can be observed only indirectly through the incomplete data set $\mathbf{g}$. A many-to-one mapping from the complete data space to the incomplete data space exists for emission tomography: given $\mathbf{c}$, there is only one particular set of projection data $\mathbf{g}$, but given $\mathbf{g}$, there are many possible realizations of $\mathbf{c}$. We have

$$G_{t\theta} = \sum_{ij} C_{t\theta;ij}.$$

The corresponding maximum likelihood estimate for the complete/incomplete data formulation can be written

$$\begin{aligned}
&\arg\max_{\mathbf{f}}\ \log \Pr(\mathbf{G}=\mathbf{g}|\mathbf{f}) \\
&= \arg\max_{\mathbf{f}}\ \log \frac{\Pr(\mathbf{G}=\mathbf{g}|\mathbf{C}=\mathbf{c},\mathbf{f})\Pr(\mathbf{C}=\mathbf{c}|\mathbf{f})}{\Pr(\mathbf{C}=\mathbf{c}|\mathbf{G}=\mathbf{g},\mathbf{f})} \\
&= \arg\max_{\mathbf{f}}\ [\log \Pr(\mathbf{G}=\mathbf{g}|\mathbf{C}=\mathbf{c},\mathbf{f}) + \log \Pr(\mathbf{C}=\mathbf{c}|\mathbf{f}) \\
&\qquad - \log \Pr(\mathbf{C}=\mathbf{c}|\mathbf{G}=\mathbf{g},\mathbf{f})] \qquad (3.11) \\
&= \arg\max_{\mathbf{f}}\ [\log \Pr(\mathbf{C}=\mathbf{c}|\mathbf{f}) - \log \Pr(\mathbf{C}=\mathbf{c}|\mathbf{G}=\mathbf{g},\mathbf{f})]\,. \qquad (3.12)
\end{aligned}$$

Because of the many-to-one mapping from the complete data space, $\mathbf{C}$, to the incomplete data space $\mathbf{G}$, the value of $\Pr(\mathbf{G}=\mathbf{g}|\mathbf{C}=\mathbf{c},\mathbf{f})$ in Equation 3.11 becomes unity. This makes the corresponding $\log(\cdot)$ term go to zero. Now, taking the conditional expectation with respect to the complete data field $\mathbf{C}$ of both sides, conditioned on a current estimate, $\hat{\mathbf{f}}^n$, and projection data, $\mathbf{g}$, (we use the notation of Table 3.1, $E_{\mathbf{X}}\{\cdot\}$, to denote the expected value over a random field X), Equation 3.12 becomes

$$\arg\max_{\mathbf{f}}\ \log \Pr(\mathbf{G}=\mathbf{g}|\mathbf{f}) = \arg\max_{\mathbf{f}}\ \left[Q(\mathbf{f}|\hat{\mathbf{f}}^n) - H(\mathbf{f}|\hat{\mathbf{f}}^n)\right].$$

Since the left side of the equation does not involve $\mathbf{C}$, it is unchanged. Here,

$$Q(\mathbf{f}|\hat{\mathbf{f}}^n) \overset{\text{def}}{=} \sum_{\{\mathbf{c}\}}[\log \Pr(\mathbf{C}=\mathbf{c}|\mathbf{f})]\Pr(\mathbf{C}=\mathbf{c}|\mathbf{G}=\mathbf{g},\hat{\mathbf{f}}^n)$$

$$= \quad E_{\mathbf{C}}\{\log \Pr(\mathbf{C}|\mathbf{f})|\mathbf{G} = \mathbf{g}, \hat{\mathbf{f}}^n\},$$

and

$$\begin{aligned} H(\mathbf{f}|\hat{\mathbf{f}}^n) &\stackrel{\text{def}}{=} \sum_{\{\mathbf{c}\}} [\log \Pr(\mathbf{C} = \mathbf{c}|\mathbf{G} = \mathbf{g}, \mathbf{f})] \Pr(\mathbf{C} = \mathbf{c}|\mathbf{G} = \mathbf{g}, \hat{\mathbf{f}}^n) \\ &= E_{\mathbf{C}}\{\log \Pr(\mathbf{C}|\mathbf{G} = \mathbf{g}, \mathbf{f})|\mathbf{G} = \mathbf{g}, \hat{\mathbf{f}}^n\}. \end{aligned}$$

According to Jensen's inequality [47], any update of $\mathbf{f}$, relative to the current estimate $\hat{\mathbf{f}}^n$, is guaranteed to decrease the value of $H(\mathbf{f}|\hat{\mathbf{f}}^n)$. Hence,

$$H(\mathbf{f}^{n+1}|\mathbf{f}^n) \leq H(\mathbf{f}^n|\mathbf{f}^n) \tag{3.13}$$

for any $\mathbf{f}^{n+1} \neq \mathbf{f}^n$. Note that $E_{\mathbf{X}}\{\mathbf{X}|\mathbf{Y} = \mathbf{y}\}$ denotes an expectation of a random field $\mathbf{X}$ conditioned on $\mathbf{Y}$, $E_{\mathbf{X}}\{\Pr(\mathbf{X} = \mathbf{x}|\mathbf{Y} = \mathbf{y})\}$ denotes an expectation of the conditional probability function of a random field $\mathbf{X}$, and $E_{\mathbf{X}}\{\Pr(\mathbf{X} = \mathbf{x}|\mathbf{Y} = \mathbf{y})|\mathbf{Y} = \mathbf{y}\}$ denotes an conditional expectation of the conditional probability function of a random field $\mathbf{X}$.

Maximizing the log-likelihood, therefore, reduces to maximizing $Q(\mathbf{f}|\hat{\mathbf{f}}^n)$. The EM algorithm for the ML estimate is an iterative scheme where each iteration consists of the following two steps.

E-step : Form $\quad Q(\mathbf{f}|\hat{\mathbf{f}}^n) = E_{\mathbf{C}}\{\log \Pr(\mathbf{C}|\mathbf{f})|\mathbf{G} = \mathbf{g}, \hat{\mathbf{f}}^n\}$

M-step : Solve $\quad n \rightarrow n + 1,$

$$\hat{\mathbf{f}}^{n+1} = \arg\max_{\mathbf{f}} \; Q(\mathbf{f}|\hat{\mathbf{f}}^n)$$

E-step: The complete data $C_{t\theta;ij}$ is modeled as independent Poisson random variables with mean values $\mathcal{H}_{t\theta;ij} f_{ij}$, where f_{ij} is the source intensity at (i, j). Hence the conditional probability,

$$\Pr(\mathbf{C}|\mathbf{f}) = \prod_{t\theta} \prod_{ij} \frac{(\mathcal{H}_{t\theta;ij} f_{ij})^{C_{t\theta;ij}} e^{-\mathcal{H}_{t\theta;ij} f_{ij}}}{C_{t\theta;ij}!}.$$

Taking the logarithm of the both sides of the above equation yields

$$\begin{aligned} \log \Pr(\mathbf{C}|\mathbf{f}) &= \sum_{t\theta} \sum_{ij} [C_{t\theta;ij} \log(\mathcal{H}_{t\theta;ij} f_{ij}) + (-\mathcal{H}_{t\theta;ij} f_{ij}) - \log(C_{t\theta;ij}!)] \\ &= \sum_{t\theta} \sum_{ij} [C_{t\theta;ij} \log f_{ij} + (-\mathcal{H}_{t\theta;ij} f_{ij})] + \text{ terms independent of } \mathbf{f}. \end{aligned}$$

Now, taking the conditional expectation with respect to the complete data space of the above equation, we obtain,

$$\begin{aligned} &E_{\mathbf{C}}\{\log \Pr(\mathbf{C}|\mathbf{f}) | \mathbf{G} = \mathbf{g}, \hat{\mathbf{f}}^n\} \\ &\quad = \sum_{t\theta} \sum_{ij} \left[E_{\mathbf{C}}\{C_{t\theta;ij} | \mathbf{G} = \mathbf{g}, \hat{\mathbf{f}}^n\} \log f_{ij} + (-\mathcal{H}_{t\theta;ij} f_{ij}) \right] \\ &\qquad\qquad + \text{ terms independent of } \mathbf{f}. \end{aligned} \tag{3.14}$$

We, therefore, need only to evaluate the term $E_{\mathbf{C}}\{C_{t\theta;ij}|\mathbf{G}=\mathbf{g},\hat{\mathbf{f}}^n\}$ for the E-step. As is shown in Appendix A [20, 50],

$$E_{\mathbf{C}}\{C_{t\theta;ij}|\mathbf{G}=\mathbf{g},\hat{\mathbf{f}}^n\} = g_{t\theta}\frac{\mathcal{H}_{t\theta;ij}\hat{f}^n_{ij}}{\sum_{kl}\mathcal{H}_{t\theta;kl}\hat{f}^n_{kl}}. \tag{3.15}$$

Now, the E-step can be written

$$\begin{aligned} Q(\mathbf{f}|\hat{\mathbf{f}}^n) &= E_{\mathbf{C}}\{\log\Pr(\mathbf{C}|\mathbf{f})|\mathbf{G}=\mathbf{g},\hat{\mathbf{f}}^n\} \\ &= \sum_{ij}\left[\left(-\sum_{t\theta}\mathcal{H}_{t\theta,ij}\right)f_{ij} + \sum_{t\theta}\frac{g_{t\theta}\mathcal{H}_{t\theta,ij}\hat{f}^n_{ij}}{\sum_{kl}\mathcal{H}_{t\theta,kl}\hat{f}^n_{kl}}\log f_{ij}\right] \\ &\quad + \text{terms independent of } \mathbf{f}. \end{aligned} \tag{3.16}$$

We show in Appendix B that the above objective function in Equation 3.16 can also be obtained by using an algebraic transformation of the Poisson log-likelihood function in Equation 3.10.

M-step: It is easy to show that the Poisson log-likelihood is concave. Recall the log-likelihood function $\log\Pr(\mathbf{G}=\mathbf{g}|\mathbf{f})$ in Equation 3.9. If we take a second partial derivative of $\log\Pr(\mathbf{G}=\mathbf{g}|\mathbf{f})$, we get

$$\frac{\partial^2\log\Pr(\mathbf{G}=\mathbf{g}|\mathbf{f})}{\partial f_{kl}\partial f_{mn}} = -\sum_{t\theta} g_{t\theta}\frac{\mathcal{H}_{t\theta;kl}\mathcal{H}_{t\theta;mn}}{(\sum_{ij}\mathcal{H}_{t\theta;ij}f_{ij})^2}.$$

which is always less than or equal to zero for all $k \neq m$ and $l \neq n$. For any $\mathbf{f}$,

$$\sum_{kl}\sum_{mn} f_{kl} \frac{\partial^2 \log \Pr(\mathbf{G} = \mathbf{g}|\mathbf{f})}{\partial f_{kl} \partial f_{mn}} f_{mn}$$

is always less than or equal to zero. The above equation becomes zero when $g_{t\theta} = 0$ for all (t, θ). The log-likelihood, hence, is concave.

Now the maximization step becomes trivial - we simply take partial derivatives and set them equal to zero:

$$\frac{\partial Q(\mathbf{f}|\hat{\mathbf{f}}^n)}{\partial f_{ij}} = -\sum_{t\theta} \mathcal{H}_{t\theta;ij} + \sum_{t\theta} \frac{g_{t\theta}\mathcal{H}_{t\theta;ij}\hat{f}_{ij}^n}{\sum_{kl} \mathcal{H}_{t\theta;kl}\hat{f}_{kl}^n} \frac{1}{f_{ij}} = 0. \tag{3.17}$$

This, then, gives an update equation for f_{ij},

$$\hat{f}_{ij}^{n+1} = \frac{\sum_{t\theta} \frac{g_{t\theta}\mathcal{H}_{t\theta,ij}\hat{f}_{ij}^n}{\sum_{kl} \mathcal{H}_{t\theta,kl}\hat{f}_{kl}^n}}{\sum_{t\theta} \mathcal{H}_{t\theta,ij}}, \tag{3.18}$$

with an initial guess $\hat{f}_{ij}^0 > 0$ for all (i, j). Thus, the EM-ML approach becomes an explicit iterative reconstruction procedure. Notice that the intensity values are always nonnegative at each iteration. Since $\mathcal{H}_{t\theta;ij}$ is the probability of a photon detected at (t, θ) being emitted from (i, j), $\sum_{t\theta} \mathcal{H}_{t\theta;ij} = 1$. By summing over all pixels (i, j) of Equation 3.18, it is easy to show that total photon count is preserved:

$$\sum_{ij} \hat{f}_{ij}^{n+1} = \sum_{t\theta} g_{t\theta}.$$

Due, however, to the inherently ill-posed nature of the unconstrained ML estimation (see page 100 in [54]), the reconstructed image deteriorates with increasing iterations [56]. This effect often called the "checkerboard effect" has been addressed by Snyder *et al.* [52]. Several methods have been suggested to overcome this effect. Veklerov and Llacer [35], for instance, proposed stopping rules in the iteration process. Snyder and Miller [51] proposed the use of sieves to smooth the reconstruction.

A different approach taken by several authors [16, 19, 20, 21, 29, 33, 34] is to model the reconstruction problem using the Bayesian paradigm, which includes prior knowledge of **F**. In the Bayesian approach, **F** is now modeled as a random field. A prior can then be assumed. The approach is successful in overcoming the checkerboard effect and also has better convergence properties than the ML-EM algorithm. The majority of the work described in this thesis focuses on how to model the prior information to best capture our knowledge.

3.2.2 Bayesian Approach

By Bayes Theorem, the *a posteriori* conditional probability density of **F** given the projection measurements **G** is

$$\Pr(\mathbf{F}=\mathbf{f}|\mathbf{G}=\mathbf{g}) = \frac{\Pr(\mathbf{G}=\mathbf{g}|\mathbf{F}=\mathbf{f})\Pr(\mathbf{F}=\mathbf{f})}{\Pr(\mathbf{G}=\mathbf{g})}, \tag{3.19}$$

where $\Pr(\mathbf{G}=\mathbf{g}|\mathbf{F}=\mathbf{f})$ is the likelihood function and, $\Pr(\mathbf{F}=\mathbf{f})$ is the *a priori* probability distribution of **F**. Note that $\Pr(\mathbf{G}=\mathbf{g})$ is a constant term independent of $\mathbf{f}$. (The probability of the random field $\mathbf{G}$ having particular measurements $\mathbf{g}$ is some number.) There are several ways to approach the above Bayesian problem [44]. We can use a minimum mean squared estimate (MMSE) to obtain an estimate of the expected value of $\mathbf{F}$ conditioned on $\mathbf{G}$,

$$\hat{\mathbf{f}}_{\text{MMSE}} = E_{\mathbf{F}}\{\mathbf{F}|\mathbf{G} = \mathbf{g}\},$$

where the notation $E_{\mathbf{X}}\{\cdot\}$ denotes the expectation over a random field $\mathbf{X}$.

We can also find the maximum *a posteriori* (MAP) estimate of $\mathbf{f}$ by maximizing $\Pr(\mathbf{F} = \mathbf{f}|\mathbf{G} = \mathbf{g})$,

$$\hat{\mathbf{f}}_{\text{MAP}} = \arg\max_{\mathbf{f}} \; \Pr(\mathbf{F} = \mathbf{f}|\mathbf{G} = \mathbf{g})$$

If the posterior distribution, $\Pr(\mathbf{F} = \mathbf{f}|\mathbf{G} = \mathbf{g})$, is symmetric and unimodal, then $\hat{\mathbf{f}}_{\text{MMSE}} = \hat{\mathbf{f}}_{\text{MAP}}$. Also note that the ML estimate can be considered a special case of the MAP estimate when the prior distribution, $\Pr(\mathbf{F} = \mathbf{f})$, is uniform.

The MAP estimate is the method that we are interested in. Using the logarithm (monotonic) function, we rewrite the MAP estimate as

$$\begin{aligned}
&\arg\max_{\mathbf{f}} \; \log\Pr(\mathbf{F} = \mathbf{f}|\mathbf{G} = \mathbf{g}) \\
&= \quad \arg\max_{\mathbf{f}} \; [\log\Pr(\mathbf{G} = \mathbf{g}|\mathbf{F} = \mathbf{f}) + \log\Pr(\mathbf{F} = \mathbf{f}) - \log\Pr(\mathbf{G} = \mathbf{g})] \\
&= \quad \arg\max_{\mathbf{f}} \; [\log\Pr(\mathbf{G} = \mathbf{g}|\mathbf{F} = \mathbf{f}) + \log\Pr(\mathbf{F} = \mathbf{f})]\,. \qquad (3.20)
\end{aligned}$$

One approach for the MAP estimate is to again use a complete/incomplete data formulation. The estimate can then be written

$$\begin{aligned}
&\arg\max_{\mathbf{f}}\ [\,\log\Pr(\mathbf{G}=\mathbf{g}|\mathbf{F}=\mathbf{f})+\log\Pr(\mathbf{F}=\mathbf{f})\,] \\
&= \arg\max_{\mathbf{f}}\ [\,\log\Pr(\mathbf{C}=\mathbf{c}|\mathbf{F}=\mathbf{f})-\log\Pr(\mathbf{C}=\mathbf{c}|\mathbf{G}=\mathbf{g},\mathbf{F}=\mathbf{f}) \\
&\qquad +\log\Pr(\mathbf{F}=\mathbf{f})\,]. \qquad (3.21)
\end{aligned}$$

As with the EM-ML approach, we take the conditional expectation (with respect to the complete data space $\mathbf{C}$) of both sides in Equation 3.21, conditioned on a current estimate, $\hat{\mathbf{f}}^n$, and projection data, $\mathbf{g}$. Since the left side of the equation does not involve $\mathbf{C}$, it is unchanged, as is the term involving $\log\Pr(\mathbf{F}=\mathbf{f})$. Now, we obtain

$$\begin{aligned}
&\arg\max_{\mathbf{f}}\ \log\Pr(\mathbf{F}=\mathbf{f}|\mathbf{G}=\mathbf{g}) \\
&= \arg\max_{\mathbf{f}}\ \left[Q(\mathbf{f}|\hat{\mathbf{f}}^n)-H(\mathbf{f}|\hat{\mathbf{f}}^n)+\log\Pr(\mathbf{F}=\mathbf{f})\right], \\
&= \arg\max_{\mathbf{f}}\ B(\mathbf{f}|\mathbf{g}), \qquad (3.22)
\end{aligned}$$

where

$$B(\mathbf{f}|\mathbf{g}) \overset{\text{def}}{=} Q(\mathbf{f}|\hat{\mathbf{f}}^n)-H(\mathbf{f}|\hat{\mathbf{f}}^n)+\log\Pr(\mathbf{F}=\mathbf{f}).$$

As mentioned previously, $H(\mathbf{f}|\hat{\mathbf{f}}^n)$ decreases at every iteration. The EM-MAP algorithm, hence, consists of the following two steps:

$$\text{E-step : Form} \quad Q(\mathbf{F}|\hat{\mathbf{f}}^n) = E_{\mathbf{c}}\{\log \Pr(\mathbf{C}|\mathbf{F}=\mathbf{f})|\mathbf{G}=\mathbf{g}, \mathbf{F}=\hat{\mathbf{f}}^n\} \tag{3.23}$$

$$\text{M-step : Solve} \quad \arg\max_{\mathbf{f}} \left[Q(\mathbf{f}|\hat{\mathbf{f}}^n) + \log \Pr(\mathbf{F}=\mathbf{f})\right] \tag{3.24}$$

Including the prior term may, however, make the M-step function non-concave. In general, finding a global maximum of the M-step can become a difficult problem in this case. We can instead employ the generalized EM method.

3.2.2.1 Generalized EM (GEM)

In the GEM approach, the M-step is replaced by finding an $\mathbf{f}$ that increases the M-step function, instead of one that maximizes the function at each iteration. We write the GEM procedure

$$\begin{aligned} &\text{E-step : Form} && E_{\mathbf{C}}\{\log \Pr(\mathbf{C}|\mathbf{F}=\mathbf{f})|\mathbf{G}=\mathbf{g}, \mathbf{F}=\hat{\mathbf{f}}^n\} \\ &\text{M-step : Find} && \mathbf{f} \text{ that satisfies} \\ & && [Q(\mathbf{f}|\hat{\mathbf{f}}^n) + \log \Pr(\mathbf{F}=\mathbf{f})] > [Q(\hat{\mathbf{f}}^n|\hat{\mathbf{f}}^n) + \log \Pr(\mathbf{F}=\hat{\mathbf{f}}^n)]. \end{aligned}$$

Notice that the EM method is a special case of Generalized EM in which we search for $\mathbf{f}$ that maximizes $Q(\mathbf{f}|\hat{\mathbf{f}}^n) + \log \Pr(\mathbf{F}=\mathbf{f})$. Following from the definition of GEM and the inequality of Equation 3.13,

$$B(\hat{\mathbf{f}}^{n+1}|\mathbf{g}) \;\geq\; B(\hat{\mathbf{f}}^{n}|\mathbf{g}),$$

for any sequence $\{\hat{\mathbf{f}}\}$ of a GEM algorithm. Hebert and Leahy [20] employed the GEM approach in emission tomographic reconstruction using a prior based on a Gibbs distribution.

Here, we briefly discuss the general convergence of EM/GEM approaches [20]. Since both EM and GEM approaches ensure an increase in $B(\mathbf{f}|\mathbf{g})$ at every M-step, their solutions converge monotonically to some B^*, if $B(\mathbf{f}|\mathbf{g})$ is bounded from above. For the EM approach, even if we perform a global maximization in the M-step, there is no guarantee that B^* is the global maximum of $B(\mathbf{f}|\mathbf{g})$ due to the $H(\mathbf{f}|\hat{\mathbf{f}}^n)$ term: the M-step only involves maximization with respect to $Q(\mathbf{f}|\hat{\mathbf{f}}^n)$, which is only a part of the posterior distribution. The remainder of the distribution, $-H$, is increased but not maximized. In general, the EM/GEM approach only assures that the sequence $\{\hat{\mathbf{f}}^n\}$ will converge to a fixed point of $B(\mathbf{f}|\mathbf{g})$, given that $Q(\mathbf{f}|\hat{\mathbf{f}}^n)+\log\Pr(\mathbf{F}=\mathbf{f})$ is continuous with respect to both $\mathbf{f}$ and $\hat{\mathbf{f}}^n$ [58]. The convergence of the EM–ML algorithm of our particular case, a Poisson likelihood distribution, is shown in Appendix B. We show that the EM-ML update equation of a Poisson likelihood distribution can be obtained by using an algebraic transformation of the Poisson log-likelihood function followed by a coordinate-wise optimization.

Chapter 4

Incorporating Intermodality Information in the Prior

In both functional imaging PET and SPECT, there is a significant loss of information in going from object to projection data due to the projection imaging process, resulting in an image of poor quality. In an attempt to improve reconstructions, several authors [16, 19, 20, 21, 29, 33, 34] have proposed probabilistic frameworks incorporating object priors in their models. Most of these models were formulated with the assumption that the object being imaged is generally smooth with some discontinuities (corresponding to boundaries). These assumptions are generic in that they apply to the general class of objects under consideration. One might well

---

The work in this chapter reflects joint work done with my colleagues Gene Gindi, Anand Rangarajan, and George Zubal. The description here follow closely the written account in [18].

suppose that prior information concerning the smoothness and discontinuities of the *specific* object being imaged should provide some additional advantage relative to generic information.

An opportunity to obtain such prior information is provided by recent results in the registration of anatomical CT or MR images with functional (SPECT or PET) images from the same patient [8, 17]. More recently, these registration techniques have been applied to brain SPECT perfusion images for registration with MRI anatomy [59, 62]. In general, CT and MR images provide high spatial resolution anatomical information. We hope to obtain improved reconstruction by integrating the images of multiple modalities. To the extent that boundaries of some anatomical objects correspond to edges in the spatial distribution of radiopharmaceuticals, we indeed get the nongeneric prior information (intermodality information) from registered CT/MR images. Previous work incorporating anatomical priors has been reported in [7, 30].

4.1 Anatomical Image as Prior

Whether the edge information obtained from anatomical images can serve as a useful prior for functional image reconstruction depends on the degree of spatial correlation of anatomical structure with the functional radiopharmaceutical distribution. In the cases where the correlation is good, we may expect the anatomy to serve as a useful prior.

One example of good anatomical and functional correlation can be seen in the autoradiograph in Figure 4.1, obtained from a study [57] involving radiopharmaceuticals for brain perfusion imaging. The autoradiograph is obtained from thin physical slices of a monkey brain, after injecting a radiopharmaceutical which accumulates in tissue corresponding to the blood flow. The autoradiograph shows a strong correlation of the blood flow distribution to the anatomy. The boundary between grey (the dark convoluted outer region) and white matter (the light intermediate region) is distinct, as is the boundary between white matter and the internal structure (the dark inner region). Since these and other regions can be well delineated on anatomical images, it may be possible that selected boundaries obtained from CT or MRI can be used as nongeneric (intermodality) prior in a functional image reconstruction. The autoradiograph also shows a reasonable uniformity of the distribution within each region, supporting the generic smoothness assumption.

In this chapter, the prior information obtained from the anatomical image is represented by an "edge map" $\mathbf{e}$ with elements e_{ij} where $0 \leq e_{ij} \leq 1$. A value of $e_{ij} = 1$ represents the highest level of confidence that a corresponding functional edge exists at location (i, j). We have two edge maps: $\mathbf{e}^h$ for horizontal edges and $\mathbf{e}^v$ for vertical edges. For brevity we shall use $\mathbf{e}$ (or e_{ij}) without superscript, except where necessary. A diagonal edge can be approximated as a combination of vertical and horizontal edges following a staircase shape. The edge map is to be incorporated in the reconstruction model as prior information and the reconstruction model should have the following two properties: i) the edge map should not overrule the data – an edge in the functional image should be reconstructed even if it is not

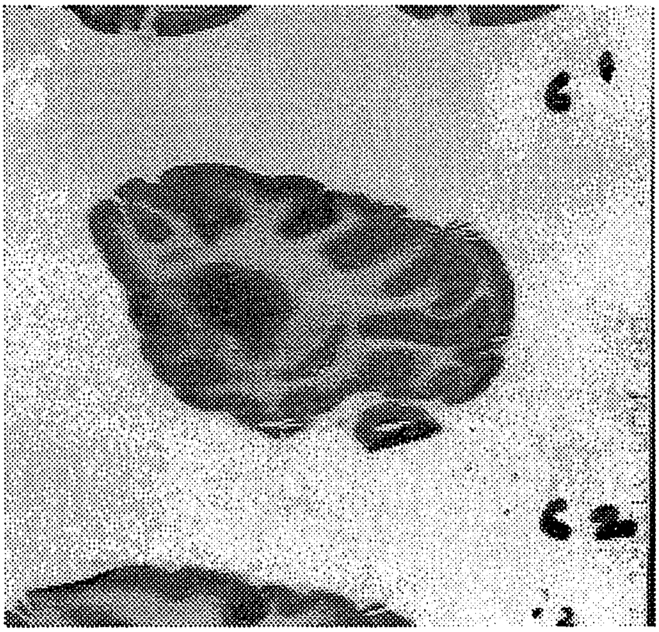


Figure 4.1: Autoradiograph of a monkey brain

promoted by the edge map, and ii) a strong edge map value should not create a false edge in the functional image.

The values of the edge map are not derived by simply running an edge operator over the anatomical image since the value of e_{ij} should depend on the *particular* anatomical edge under consideration. In order to set e_{ij} to a high value, we need both a strong anatomical edge and a reasonable expectation that there exists a corresponding edge in the functional image at location (i, j). A principled way of deriving the edge map may involve an intelligent segmentation procedure that can correctly label all anatomical regions. Values of the edge map can then be assigned based on the locations of the edges, the tissue types of the adjacent regions, and the type of radiopharmaceutical used in the study. The correlation of functional and anatomical edges is likely to depend on all three of these factors. Given the edge map from the anatomical image and given the registration transformation as derived from one of several applicable techniques, we then have enough information to apply e_{ij}. For our simulations, the edge map is derived by hand tracing only those anatomical edges that are likely to correlate with functional edges and assigning values to those edges.

In practice, the correlation between anatomical and functional edges involves some level of position resolution error. Here, we classify such errors into three types:

1) An anatomical edge may be grossly misaligned by many pixels relative to the functional edge, completely missing (for example, in the case of abnormal concentration of radioactivity corresponding to a diseased area within an organ, the diseased area may not be seen in the anatomical image), or present as an extra, inappropriate edge that has no corresponding functional edge.

2) An anatomical edge may be misaligned up to a few pixel distances in the functional image due to intermodality registration error.

3) An anatomical edge may not be localized within the pixel resolution of the functional image reconstruction. Because a typical slice thickness in the functional image acquisition is about 0.7cm – 1cm, a three-dimensional boundary in the distribution is reconstructed as a blurred two-dimensional boundary averaging over the slice thickness. This is often referred as the partial volume effect. In addition, due to the slice thickness of the anatomical image itself (0.5mm – 1.5mm in CT/MR), averaging also occurs in anatomical images and there may not exist a

highly localized boundary in the anatomical images to start with. For some region boundaries, this transition from one tissue type to another may occur over several resolution distances in the anatomical image.

In our work, the effects of 1) are modelled by including extra and missing edges in the simulations. The effects of 3) are crudely modelled by blurring the anatomical edges resulting a thin band rather than a thin line. A proper model for 2) might include random misalignments, by a few pixels, of sections of anatomical boundary. A detailed study would include an assessment of the effects on the reconstructions of shifting an anatomical edge in a controlled manner away from its nominal position. In the initial work presented here, we lump the effects of 2) and 3) into our blur model.

4.2 A Model for Bayesian Reconstruction with Intermodality Information

We now formulate a Bayesian model for our reconstruction problem. In a Bayesian model, the posterior distribution is obtained from the likelihood and prior distributions (Section 3.2.2). The prior represents our knowledge of the spatial distribution of the source. Here, we model the likelihood as a Poisson distribution (Equation 3.8) as discussed in Section 3.2.1, and the prior as a Gibbs distribution for the reasons to be discussed in Section 4.2.2.2.

The reconstruction problem is to estimate a 2-D source distribution $\mathbf{f}$ given projection data $\mathbf{g}$ of the source and a registered edge map $\mathbf{e}$, which is a deterministic

Table 4.1: Symbols used in Chapter 4

$\mathbf{e}^h, \mathbf{e}^v$	Horizontal, vertical anatomical edge confidence maps
λ_1	Prior weighting parameter
κ_1, κ_2	Energy function parameters
α^h_{ij}	Space variant weak membrane parameter $\kappa_1(1 - e^h_{ij}) + \kappa_2 e^h_{ij}$
α^v_{ij}	Space variant weak membrane parameter $\kappa_1(1 - e^v_{ij}) + \kappa_2 e^v_{ij}$
$\phi^*(\cdot)$	Prior potential function obtained by eliminating line processes: Broken parabola
$\phi^{GM}(\cdot), \phi^{GR}(\cdot)$	Prior potential functions that approximate $\phi^*(\cdot)$
$E_{P1}(\mathbf{f}, \mathbf{l})$	Energy function for the prior on $\mathbf{f}$ and $\mathbf{l}$
$E_1(\mathbf{f}, \mathbf{l})$	Overall objective function
$E_1^*(\mathbf{f})$	Equivalent overall objective function obtained by eliminating line processes
$E_1^{GR}(\mathbf{f})$	Approximated objective function used in optimization

quantity obtained from an anatomical modality. The estimate should be consistent with the projection data and be generally smooth except at locations of significant edges in the source distribution $\mathbf{f}$. Locating these edges is aided by high values of $\mathbf{e}$. Table 4.1 summarizes our notational definitions in this chapter and Table 3.1 shows the notational definitions used throughout this thesis.

4.2.1 Formulation with Line Processes

We first formulate our reconstruction problem with the line processes as proposed by Geman and Geman [15] to model discontinuities in the 2-D image **f**, though we shall eventually dispense with the line processes. We model the prior via the use of binary-valued line processes which are unobservables corresponding to image discontinuities. At locations where the line processes are "off", the prior encourages smoothness by penalizing the formation of high image-intensity gradients; at locations where the line processes are "on", this penalty is removed and high intensity gradients are permitted. The prior consists of both continuous variables (the source intensities) and binary variables (the line processes), and we are interested in simultaneously estimating them. A value $l^h_{ij} = 1$ corresponds to the presence of a horizontal line process located between pixel locations $(i+1, j)$ and (i, j). A value $l^v_{ij} = 1$ corresponds to the presence of a vertical line process located between pixel locations (i, j) and $(i, j+1)$. Locations of both l^v_{ij} and l^h_{ij} are shown in Figure 4.2. For brevity, we shall use the quantity **l** with no superscripts except when necessary.

We may now write Bayes theorem:

$$\Pr(\mathbf{F} = \mathbf{f}, \mathbf{L} = \mathbf{l} | \mathbf{G} = \mathbf{g}) = \frac{\Pr(\mathbf{G} = \mathbf{g} | \mathbf{F} = \mathbf{f}, \mathbf{L} = \mathbf{l}) \Pr(\mathbf{F} = \mathbf{f}, \mathbf{L} = \mathbf{l})}{\Pr(\mathbf{G} = \mathbf{g})}. \tag{4.1}$$

Given the posterior distribution in Equation 4.1, one possible estimator is the max-

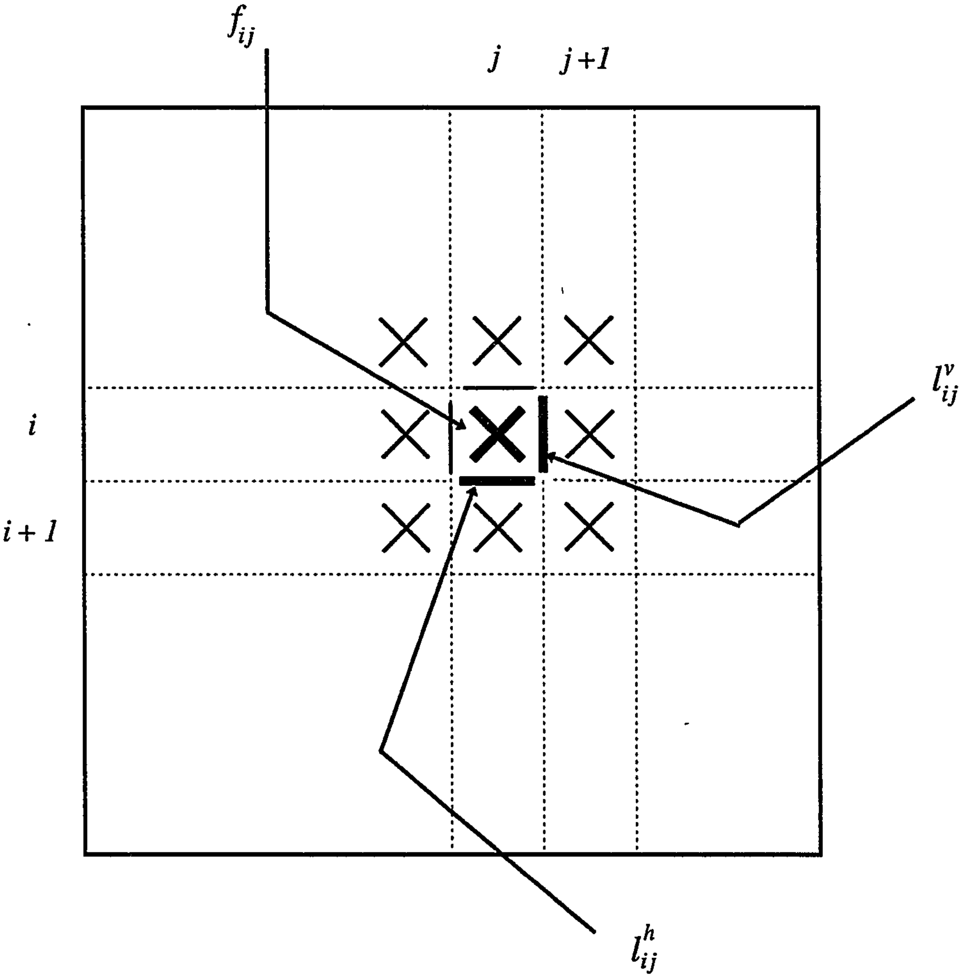


Figure 4.2: Locations of l^v_{ij} and l^h_{ij}.

imum *a posteriori* estimate (MAP). We require the following MAP estimate:

$$\begin{aligned}(\hat{\mathbf{f}}_{\mathrm{MAP}}, \hat{\mathbf{l}}_{\mathrm{MAP}}) &= \arg\max_{(\mathbf{f},\mathbf{l})} \Pr(\mathbf{F} = \mathbf{f}, \mathbf{L} = \mathbf{l} | \mathbf{G} = \mathbf{g}) && (4.2)\\ &= \arg\max_{(\mathbf{f},\mathbf{l})} \Pr(\mathbf{G} = \mathbf{g} | \mathbf{F} = \mathbf{f}, \mathbf{L} = \mathbf{l}) \Pr(\mathbf{F} = \mathbf{f}, \mathbf{L} = \mathbf{l}) && (4.3)\\ &= \arg\max_{(\mathbf{f},\mathbf{l})} \Pr(\mathbf{G} = \mathbf{g} | \mathbf{F} = \mathbf{f}) \Pr(\mathbf{F} = \mathbf{f} | \mathbf{L} = \mathbf{l}) \Pr(\mathbf{L} = \mathbf{l}). && (4.4)\end{aligned}$$

The term $\Pr(\mathbf{G} = \mathbf{g})$ can be dropped in the above equations since the arg max in Equation 4.2 is carried out only with respect to $\mathbf{f}$ and $\mathbf{l}$. Since the projection data $\mathbf{g}$ is obtained from a forward projection of the source $\mathbf{f}$, it does not involve the line processes $\mathbf{l}$. Therefore, the likelihood term $\Pr(\mathbf{G} = \mathbf{g} | \mathbf{F} = \mathbf{f}, \mathbf{L} = \mathbf{l})$ in Equation 4.3 becomes $\Pr(\mathbf{G} = \mathbf{g} | \mathbf{F} = \mathbf{f})$. The prior term, $\Pr(\mathbf{F} = \mathbf{f}, \mathbf{L} = \mathbf{l})$, in Equation 4.3 is rewritten in two terms:

$$\Pr(\mathbf{F} = \mathbf{f}, \mathbf{L} = \mathbf{l}) = \Pr(\mathbf{F} = \mathbf{f} | \mathbf{L} = \mathbf{l}) \Pr(\mathbf{L} = \mathbf{l}),$$

where the conditional probability $\Pr(\mathbf{F} = \mathbf{f} | \mathbf{L} = \mathbf{l})$ is interpreted as prior on $\mathbf{f}$ conditioned on $\mathbf{l}$, and the probability $\Pr(\mathbf{L} = \mathbf{l})$ as prior on $\mathbf{l}$ alone.

4.2.2 An Objective Function for MAP Estimation

In this section, we describe how we model each term in Equation 4.4. In particular, our focus is on modeling the Gibbs prior distribution, $\Pr(\mathbf{F} = \mathbf{f}, \mathbf{L} = \mathbf{l})$, in which the intermodality information (edge map), $\mathbf{e}$, is utilized. Using the monotonic property of the logarithm function, we rewrite our MAP estimate in Equation 4.4

$$\begin{aligned}(\hat{\mathbf{f}}_{\mathrm{MAP}},\hat{\mathbf{l}}_{\mathrm{MAP}}) &= \arg\max_{(\mathbf{f},\mathbf{l})} \Pr(\mathbf{G}=\mathbf{g}|\mathbf{F}=\mathbf{f})\Pr(\mathbf{F}=\mathbf{f}|\mathbf{L}=\mathbf{l})\Pr(\mathbf{L}=\mathbf{l}) \\ &= \arg\min_{(\mathbf{f},\mathbf{l})} \,[\, -\log\Pr(\mathbf{G}=\mathbf{g}|\mathbf{F}=\mathbf{f}) - \log\Pr(\mathbf{F}=\mathbf{f}|\mathbf{L}=\mathbf{l}) \\ &\qquad\qquad -\log\Pr(\mathbf{L}=\mathbf{l})\,] \end{aligned} \tag{4.5}$$

4.2.2.1 The Likelihood

The likelihood, $\Pr(\mathbf{G}=\mathbf{g}|\mathbf{F}=\mathbf{f})$, is modeled as a Poisson distribution (Equation 3.8). The objective function associated with the likelihood is thus

$$\begin{aligned} &-\log\Pr(\mathbf{G}=\mathbf{g}|\mathbf{F}=\mathbf{f}) \\ &= \sum_{t\theta}\left[\sum_{ij}\mathcal{H}_{t\theta;ij}f_{ij} - g_{t\theta}\log(\sum_{ij}\mathcal{H}_{t\theta;ij}f_{ij})\right] + \text{ terms independent of } \mathbf{f} \\ &= D(\mathbf{f}) + \text{ terms independent of } \mathbf{f} \end{aligned} \tag{4.6}$$

where

$$D(\mathbf{f}) \stackrel{\mathrm{def}}{=} \sum_{t\theta}\left[\sum_{ij}\mathcal{H}_{t\theta;ij}f_{ij} - g_{t\theta}\log(\sum_{ij}\mathcal{H}_{t\theta;ij}f_{ij})\right].$$

4.2.2.2 The Prior

The latter two terms in Equation 4.5 constitute the prior. It is usually assumed that the structure of images is locally correlated. The property of local correlation can be described by a Gibbs distribution (see for example, in [3] on page 195) [15, 16, 20]. In general, a Gibbs distribution has the form

$$\Pr(\mathbf{X} = \mathbf{x}) = \frac{1}{Z_{\mathbf{x}}} e^{-E(\mathbf{x})}, \tag{4.7}$$

where

$$Z_{\mathbf{x}} = \sum_{\{\mathbf{x}\}} e^{-E(\mathbf{x})} \tag{4.8}$$

is called a partition function. The partition function is the sum of probabilities of all possible configurations of $\mathbf{x}$. The notation $\sum_{\{\mathbf{x}\}}$ denotes the sum over all possible configurations of $\mathbf{x}$. The term $E(\mathbf{x})$ is referred to as the related Gibbs energy function. We, thus, model the prior as a Gibbs distribution with an associated energy function $E_{P1}(\mathbf{f}, \mathbf{l})$ and a weighting parameter λ_1,

$$\begin{aligned}\Pr(\mathbf{F} = \mathbf{f}, \mathbf{L} = \mathbf{l}) &= \frac{1}{Z_1} e^{-\lambda_1 E_{P1}(\mathbf{f},\mathbf{l})} \\ &= \Pr(\mathbf{F} = \mathbf{f} | \mathbf{L} = \mathbf{l}) \Pr(\mathbf{L} = \mathbf{l}) \\ &= \frac{1}{Z_1} e^{-\lambda_1 [E_S(\mathbf{f},\mathbf{l}) + E_L(\mathbf{l})]} = \frac{1}{Z_2} e^{-\lambda_1 E_S(\mathbf{f},\mathbf{l})} \frac{1}{Z_3} e^{-\lambda_1 E_L(\mathbf{l})},\end{aligned} \tag{4.9}$$

where $\Pr(\mathbf{F} = \mathbf{f} | \mathbf{L} = \mathbf{l})$ and $\Pr(\mathbf{L} = \mathbf{l})$ are again modeled as Gibbs distributions,

$$\Pr(\mathbf{F} = \mathbf{f} | \mathbf{L} = \mathbf{l}) = \frac{1}{Z_2} e^{-\lambda_1 E_S(\mathbf{f},\mathbf{l})},$$

and

$$\Pr(\mathbf{L}=\mathbf{l}) = \frac{1}{Z_3} e^{-\lambda_1 E_L(\mathbf{l})}.$$

Note that Z_1, Z_2, and Z_3 are normalizing constant terms independent of **f** or **l**.

We then have

$$-\log \Pr(\mathbf{F}=\mathbf{f} \mid \mathbf{L}=\mathbf{l}) = \lambda_1 E_S(\mathbf{f},\mathbf{l}) \ + \ \text{terms independent of } \mathbf{f} \text{ and } \mathbf{l},$$

where we define the energy function as

$$E_S(\mathbf{f},\mathbf{l}) \overset{\text{def}}{=} \sum_{i,j} (\Delta^v f_{ij})^2 (1 - l^h_{ij}) + \sum_{i,j} (\Delta^h f_{ij})^2 (1 - l^v_{ij}). \tag{4.10}$$

Note that the energy function, $E_S(\mathbf{f},\mathbf{l})$, is defined over the intensities and the line processes. Here we take for our definitions of discrete partial derivatives $\Delta^v f_{ij} = f_{i+1,j} - f_{ij}$ and $\Delta^h f_{ij} = f_{i,j+1} - f_{ij}$. The two terms involving $\Delta^v f$ and $\Delta^h f$ encourage smoothness except where discontinuities occur ($l_{ij} = 1$).

The term, $\Pr(\mathbf{L} = \mathbf{l})$, imposes constraints on the discontinuities and in particular incorporates anatomical boundary information.

For the prior on the line process, we have

$$-\log \Pr(\mathbf{L}=\mathbf{l}) = \lambda_1 E_L(\mathbf{l}) \ + \ \text{terms independent of } \mathbf{l},$$

where

$$\begin{aligned} E_L(\mathbf{l}) &\overset{\text{def}}{=} \kappa_1 \sum_{ij} \left(l^v_{ij} + l^h_{ij} \right) - (\kappa_1 - \kappa_2) \sum_{ij} \left(l^v_{ij} e^v_{ij} + l^h_{ij} e^h_{ij} \right) \qquad 0 \leq \kappa_2 \leq \kappa_1 \\ &= E_{L_1}(\mathbf{l}) + E_{L_2}(\mathbf{l}). \end{aligned} \tag{4.11}$$

For convenience, we define

$$E_{L_1}(\mathbf{l}) \overset{\text{def}}{=} \kappa_1 \sum_{ij} \left(l^v_{ij} + l^h_{ij} \right)$$

and

$$E_{L_2}(\mathbf{l}) \overset{\text{def}}{=} -(\kappa_1 - \kappa_2) \sum_{ij} \left(l^v_{ij} e^v_{ij} + l^h_{ij} e^h_{ij} \right).$$

The term $E_{L_1}(\mathbf{l})$ penalizes the creation of discontinuities. The penalty for creating a discontinuity is proportional to the constant κ_1. In fact, the combination of the smoothness prior on the intensities, $E_S(\mathbf{f}, \mathbf{l})$, in Equation 4.10 and the first penalty term, $E_{L_1}(\mathbf{l})$, in Equation 4.11, $(E_S + E_{L_1})$, is known as the *weak membrane* prior model which is widely used in the computer vision literature [5, 13, 45]:

$$\begin{aligned} E_{WM}(\mathbf{f}, \mathbf{l}) &= E_S(\mathbf{f}, \mathbf{l}) + E_{L_1}(\mathbf{l}) \\ &= \sum_{ij} \left[(\Delta^v f_{ij})^2 (1 - l^h_{ij}) + (\Delta^h f_{ij})^2 (1 - l^v_{ij}) \right] + \kappa_1 \sum_{ij} \left(l^v_{ij} + l^h_{ij} \right) \\ &= \sum_{ij} \left[(\Delta^v f_{ij})^2 (1 - l^h_{ij}) + \kappa_1 l^h_{ij} \right] + \sum_{ij} \left[(\Delta^h f_{ij})^2 (1 - l^v_{ij}) + \kappa_1 l^v_{ij} \right]. \end{aligned} \tag{4.12}$$

The weak membrane prior, $E_{WM}(\mathbf{f},\mathbf{l})$ encourages the image $\mathbf{f}$ to be composed of smooth patches that are broken only at locations of high intensity gradient by discontinuities (line processes, $\mathbf{l}$). As κ_1 is increased, fewer discontinuities are obtained.

The main focus of the work in this chapter is to utilize edge information from the anatomical image to influence the breaking of smoothness in the reconstruction. At locations where there exist significant anatomical edges (high values of e_{ij}), the smoothness should be broken more easily. One way to accomplish this is to reduce the penalty for creating discontinuities at those locations. This effect can be seen by rewriting Equation 4.11 as

$$\begin{aligned} E_L(\mathbf{l}) &= E_{L_1}(\mathbf{l}) + E_{L_2}(\mathbf{l}) \\ &= \kappa_1 \sum_{ij} \left(l^v_{ij} + l^h_{ij} \right) - (\kappa_1 - \kappa_2) \sum_{ij} \left(l^v_{ij} e^v_{ij} + l^h_{ij} e^h_{ij} \right) \\ &= \sum_{ij} \Big[\left[\kappa_1 (1 - e^h_{ij}) + \kappa_2 e^h_{ij} \right] l^h_{ij} + \left[\kappa_1 (1 - e^v_{ij}) + \kappa_2 e^v_{ij} \right] l^v_{ij} \Big] \qquad (4.13) \end{aligned}$$

In Equation 4.13, the new penalty, $\kappa_1(1 - e_{ij}) + \kappa_2 e_{ij}$, interpolates between two fixed constants κ_1 and κ_2 using $\mathbf{e}$. Now we can write the "prior energy":

$$\begin{aligned} E_{P1}(\mathbf{f},\mathbf{l}) &= E_S(\mathbf{f},\mathbf{l}) + E_L(\mathbf{l}) \\ &= \sum_{ij} \Big[(\Delta^v f_{ij})^2 (1 - l^h_{ij}) + \left[\kappa_1 (1 - e^h_{ij}) + \kappa_2 e^h_{ij} \right] l^h_{ij} \Big] + \\ &\quad \sum_{ij} \Big[(\Delta^h f_{ij})^2 (1 - l^v_{ij}) + \left[\kappa_1 (1 - e^v_{ij}) + \kappa_2 e^v_{ij} \right] l^v_{ij} \Big] . \qquad (4.14) \end{aligned}$$

The effect of anatomical information, **e**, is to spatially modulate the creation of discontinuities in the reconstruction by varying the penalty accordingly. For example, if e^h_{ij} is close to unity, then the energy is reduced by turning on a line process $l^h_{ij} = 1$ only if $(\Delta^v f_{ij})^2 > \kappa_2$. Conversely, if e^h_{ij} is close to zero, an edge is created only if $(\Delta^v f_{ij})^2 > \kappa_1$. Since $\kappa_1 > \kappa_2$, the threshold for creating an edge is reduced at locations where we have high confidence that an edge exists (large e_{ij}). A coupling term of this form has been proposed by Gamble *et al.* [12] and Poggio *et al.* [43] in the context of data fusion for computer vision. In their work, an attempt was made to fuse edges from registered intensity, stereo, texture, color, and motion images. In principle, the parameters κ_1 and κ_2 should be estimated from a given object or the class of objects under consideration. (For instance, we know that the typical radiopharmaceutical uptake ratio in grey and white matter of brain is 4:1. This ratio may be used to estimate κ_1 and κ_2 for the edges between grey and white matter.) By comparing the prior energy $E_{P1}(\mathbf{f}, \mathbf{l})$ with the weak membrane model E_{WM} in Equation 4.12, we can see that our prior energy $E_{P1}(\mathbf{f}, \mathbf{l})$ is a modified weak membrane model

$$E_{P1}(\mathbf{f},\mathbf{l}) = \sum_{ij} \left[\, (\Delta^v f_{ij})^2 (1 - l^h_{ij}) + \alpha^h_{ij} l^h_{ij} \, \right] + \sum_{ij} \left[\, (\Delta^h f_{ij})^2 (1 - l^v_{ij}) + \alpha^v_{ij} l^v_{ij} \, \right],$$

with space variant parameters

$$\begin{aligned} \alpha^v_{ij} &\overset{\mathrm{def}}{=} \kappa_1 (1 - e^v_{ij}) + \kappa_2 e^v_{ij}, \\ \alpha^h_{ij} &\overset{\mathrm{def}}{=} \kappa_1 (1 - e^h_{ij}) + \kappa_2 e^h_{ij}, \end{aligned} \tag{4.15}$$

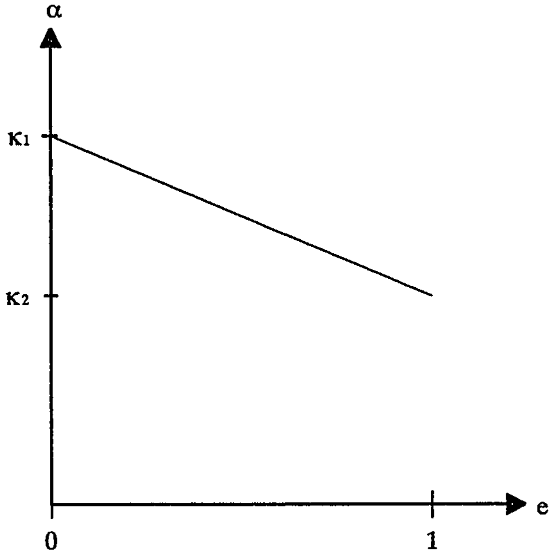


Figure 4.3: α vs. e

instead of space invariant parameter κ_1 alone. The new parameters, α_{ij}^{v} and α_{ij}^{h}, are space variant due to the space variance of $\mathbf{e}$ (shown in Figure 4.3).

Having both the likelihood and the prior defined, we now rewrite the MAP estimate

$$\begin{aligned}
(\hat{\mathbf{f}}_{\mathrm{MAP}}, \hat{\mathbf{l}}_{\mathrm{MAP}}) &= \arg\max_{(\mathbf{f},\mathbf{l})} \Pr(\mathbf{F} = \mathbf{f}, \mathbf{L} = \mathbf{l} | \mathbf{G} = \mathbf{g}) \\
&= \arg\min_{(\mathbf{f},\mathbf{l})} \left[-\log \Pr(\mathbf{F} = \mathbf{f}, \mathbf{L} = \mathbf{l} | \mathbf{G} = \mathbf{g})\right] \\
&= \arg\min_{(\mathbf{f},\mathbf{l})} \left[D(\mathbf{f}) + \lambda_1 \left(E_S(\mathbf{f},\mathbf{l}) + E_L(\mathbf{l})\right)\right] \\
&= \arg\min_{(\mathbf{f},\mathbf{l})} \left[D(\mathbf{f}) + \lambda_1 E_{P1}(\mathbf{f},\mathbf{l})\right] \qquad (4.16) \\
&= \arg\min_{(\mathbf{f},\mathbf{l})} E_1(\mathbf{f},\mathbf{l}). \qquad (4.17)
\end{aligned}$$

From the above equation, we can see that the parameter, λ_1, weights the prior with respect to the likelihood objective. We have an overall objective function

$$\begin{aligned}
E_1(\mathbf{f},\mathbf{l}) &\overset{\text{def}}{=} D(\mathbf{f}) + \lambda_1 E_{P1}(\mathbf{f},\mathbf{l}) \\
&= D(\mathbf{f}) + \lambda_1 \sum_{ij} \Big[(\Delta^v f_{ij})^2(1 - l_{ij}^h) + [\kappa_1(1 - e_{ij}^h) + \kappa_2 e_{ij}^h] l_{ij}^h\Big] \\
&\qquad +\lambda_1 \sum_{ij} \Big[(\Delta^h f_{ij})^2(1 - l_{ij}^v) + [\kappa_1(1 - e_{ij}^v) + \kappa_2 e_{ij}^v] l_{ij}^v\Big] \\
&= D(\mathbf{f}) + \lambda_1 \sum_{ij} \Big[(\Delta^v f_{ij})^2(1 - l_{ij}^h) + \alpha_{ij}^h l_{ij}^h \\
&\qquad +(\Delta^h f_{ij})^2(1 - l_{ij}^v) + \alpha_{ij}^v l_{ij}^v\Big] \qquad (4.18)
\end{aligned}$$

4.2.3 An Equivalent Objective Function

Though the objective function in Equation 4.18 embodies the problem, its minimization requires a great deal of computation. Minimizing the objective function requires a hybrid optimization technique since the intensities **f** can essentially be treated as a continuous field and subject to gradient–based optimization whereas the binary line processes **l** are subject only to combinatorial optimization. Instead of proceeding with such a hybrid optimization, we show that the binary line processes can be eliminated from the objective function leading to a new, equivalent objective function which depends only on the intensities **f**. This new objective function has the same global minima as the original one. The new objective function itself, however, needs modification due to two drawbacks which we will show below. In this section, we describe the process of modification from the original objective function in Equation 4.18 to the final objective function which we eventually utilize.

We first eliminate the line processes from the prior in Equation 4.18. Because $D(\mathbf{f})$ does not involve line processes, our minimization problem (Equation 4.16),

$$\arg\min_{(\mathbf{f},\mathbf{l})}\ [D(\mathbf{f})+\lambda_1 E_{P1}(\mathbf{f},\mathbf{l})],$$

is equivalent to

$$\arg\min_{\mathbf{f}}\ \left[D(\mathbf{f})+\lambda_1\left(\arg\min_{\mathbf{l}} E_{P1}(\mathbf{f},\mathbf{l})\right)\right].$$

Since the line processes are independent of each other, the optimization with respect to the line processes can be performed separately at each location:

$$\arg\min_{\mathbf{l}} E_{P1}(\mathbf{f},\mathbf{l})$$

becomes

$$\sum_{ij}\left[\arg\min_{l^h_{ij}\in\{0,1\}}\left[(\Delta^v f_{ij})^2(1-l^h_{ij})+\alpha^h_{ij}l^h_{ij}\right]+\arg\min_{l^v_{ij}\in\{0,1\}}\left[(\Delta^h f_{ij})^2(1-l^v_{ij})+\alpha^v_{ij}l^v_{ij}\right]\right],$$

where $l_{ij}\in\{0,1\}$ donates that the line process at location (i,j) takes a value of either 0 or 1. This reduces to a simple solution for each line process as a function of $\mathbf{f}$. Examine, for example, the term in Equation 4.14 corresponding to the horizontal line process l^h_{ij}. When $(\Delta^v f_{ij})^2\ \leq\ \alpha^h_{ij}$, the term is minimized when the line process l^h_{ij} is equal to zero. Similarly, when $(\Delta^v f_{ij})^2\ >\ \alpha^h_{ij}$, the term is minimized when

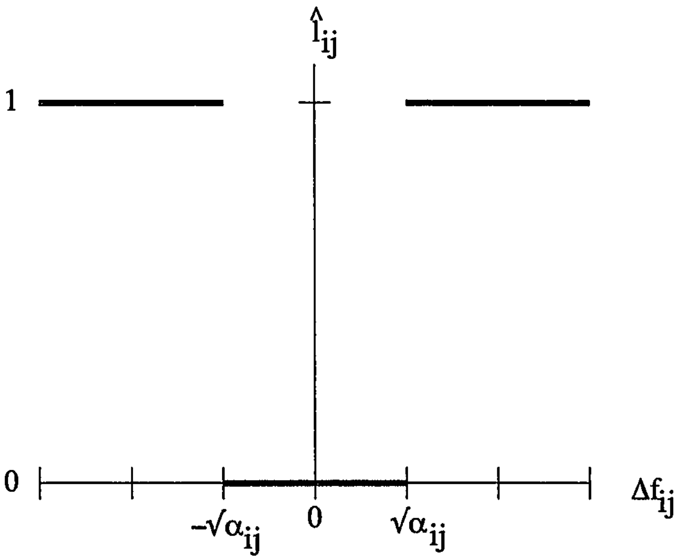


Figure 4.4: MAP solution for the line process, $\hat{l}_{ij}$.

the line process l^h_{ij} is equal to unity. The solution for the line process $\hat{l}^h_{ij}$ and $\hat{\mathrm{l}}^v_{ij}$, the MAP estimates given $\mathbf{f}$, can thus be written as

$$\hat{l}^h_{ij} = \begin{cases} 0 & (\Delta^v f_{ij})^2 \ \le \ \alpha^h_{ij} = \kappa_1(1 - e^h_{ij}) + \kappa_2 e^h_{ij} \\ 1 & (\Delta^v f_{ij})^2 \ > \ \alpha^h_{ij} = \kappa_1(1 - e^h_{ij}) + \kappa_2 e^h_{ij} \end{cases} \tag{4.19}$$

with a similar expression for the vertical line process l^v_{ij}. The MAP estimate for the line process, $\hat{l}_{ij}$, is shown in Figure 4.4.

This is a modified version of line process elimination found in Blake and Zisserman [5]. When the MAP estimate for $\mathbf{l}$ is substituted back into the objective function in Equation 4.18, we obtain an equivalent objective function which depends only on the intensities:

$$E_1^*(\mathbf{f}) = D(\mathbf{f}) + \lambda_1 \sum_{ij} \left[\phi^*(\Delta^v f_{ij}) + \phi^*(\Delta^h f_{ij})\right], \tag{4.20}$$

where the prior potential functions are

$$\phi^*(\Delta^v f_{ij}) = \begin{cases} (\Delta^v f_{ij})^2 & (\Delta^v f_{ij})^2 \le \alpha_{ij}^h \\ \alpha_{ij}^h & (\Delta^v f_{ij})^2 > \alpha_{ij}^h \end{cases} \tag{4.21}$$

with a corresponding expression for $\phi^*(\Delta^h f_{ij})$. (For the sake of convenience, we shall henceforth write expressions for ϕ^* and related functions in terms of $\Delta^v f_{ij}$ only – the dependence on $\Delta^h f_{ij}$ should then be obvious.) Figure 4.5 shows $\phi^*(\Delta f_{ij})$. Though the global minima of Equation 4.20 are the same as those of Equation 4.18 and the problems associated with binary line variables have been circumvented, there still remain problems of optimizing Equation 4.20.

4.2.4 A Modified Objective Function

There are two problems associated with using the above potential function ϕ^*. The broken parabola form of ϕ^* may make the objective function highly non-convex. (We refer to ϕ^* as a "broken parabola" due to its shape.) The combination of the likelihood term and the broken parabola prior has been analyzed for the simple case where the likelihood is of the form $||\mathbf{f} - \mathbf{g}||^2$ which corresponds to the image estimation or surface reconstruction problems in early vision work [5]. Our situation

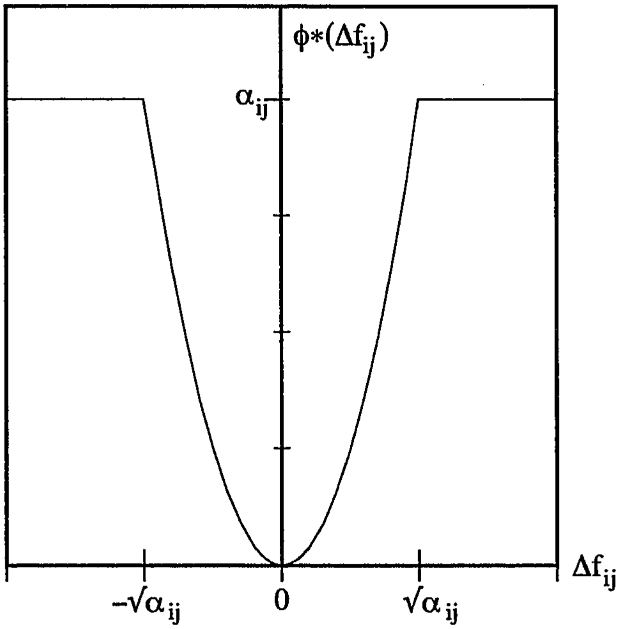


Figure 4.5: $\phi^*(\Delta f_{ij})$

in image reconstruction (Poisson likelihood) is considerably more complex despite the fact that the log-likelihood is concave. We speculate that the combination of the Poisson likelihood with the broken parabola is indeed non-convex. However, a smoother choice of such a Gibbs prior may alleviate this problem. For example, in the aforementioned early vision case, Blake and Zisserman [5] show that a smoother version of the broken parabola does make the whole objective function convex. Indeed, a number of researchers [16, 19, 20, 26] have implicitly incorporated smooth Gibbs priors of this kind into the medical reconstruction problem, but with only generic prior (no intermodality information was included).

To this end, consider the following which is a modified form of the Gibbs

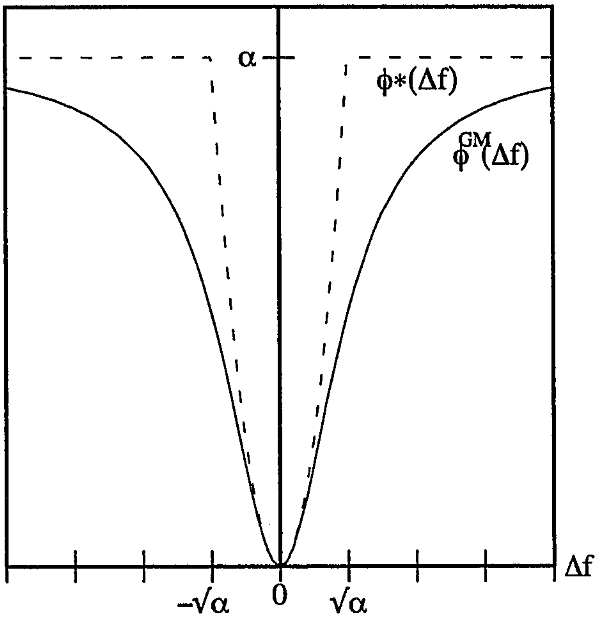


Figure 4.6: $\phi^{*}(\Delta f)$ and $\phi^{GM}(\Delta f)$

prior due to Geman and McClure [16],

$$\phi^{GM}(\Delta^{v} f_{ij}) = \frac{\alpha_{ij}^{h}(\Delta^{v} f_{ij})^{2}}{\alpha_{ij}^{h} + (\Delta^{v} f_{ij})^{2}}. \tag{4.22}$$

This Gibbs prior is displayed in Figure 4.6. This choice of ϕ^{GM} is smoother than the broken parabola and similar to smooth approximations in [5] advocated to overcome non-convexity. The heuristic transition from ϕ^{*} to ϕ^{GM} can be motivated by considering the following modification due to [45] of the original objective function in Equation 4.18. The prior, $E_{P1}(\mathbf{f},\mathbf{l})$ in Equation 4.14, is modified in the following:

$$(\Delta^v f_{ij})^2(1 - l^h_{ij}) + \alpha^h_{ij} l^h_{ij} \quad \Rightarrow \quad (\Delta^v f_{ij})^2(1 - l^h_{ij})^2 + \alpha^h_{ij}(l^h_{ij})^2$$

with the line process l^h_{ij} now assuming continuous values in the unit interval $[0,1]$. With this form of the prior, the line process can be eliminated as shown in [45] to obtain the functional form in Equation 4.22. Thus the modification from ϕ^* to ϕ^{GM} is not totally *ad hoc* and corresponds to a modification of the original prior from a binary to a continuous line process.

The second drawback, first pointed out by Geman and Reynolds [14] in the context of image restoration is characteristic of the broken parabola and still persists in the ϕ^{GM} formulation. Qualitatively speaking, the drawback is manifested as local high-intensity "hot spots" in the reconstruction. The manifestation of hot spots can be understood by examining the structure of the $\phi^{GM}(\Delta f)$ function (see Figure 4.6). Note that as Δf approaches zero, the derivative of $\phi^{GM}(\Delta f)$ approaches zero. In addition, as Δf approaches infinity, the derivative again approaches zero as seen in the flat region of the ϕ^{GM} function. In between, the derivative is positive. The degree of smoothing depends on the magnitude of the derivative of the $\phi^{GM}(\Delta f)$ function. In the two regions, close to the origin, and far from the origin (the flat region), the derivative of $\phi^{GM}(\Delta f)$ is very small. In pixel neighborhoods, where the derivatives are very small, the prior is small, and the reconstruction is driven solely by the likelihood term. The likelihood term may then create a local, high intensity, single pixel, hot spot (due to a particular noise realization) which causes the magnitudes of Δf at that location to become quite large. Ordinarily, the hot spot would be eliminated in subsequent iterations by the smoothing action of the prior. However, if the gradients at that spot (Δf) happen to be large enough, so that the flat region of the ϕ^{GM} function applies, then smoothing is again suspended and the hot spot remains. Thus the small derivatives of ϕ^{GM} at both the origin and flat segments conspire to create hot spots.

One way to overcome this conspiracy is to eliminate the vanishing derivative of ϕ^{GM} at the origin. This does not permit the likelihood to create hot spots by the above mentioned mechanism. Consider the following modification of ϕ^{GM} to a cusp-like form (see Figure 4.7(a)) where the derivative is non-vanishing at the origin:

$$\phi^{GR}(\Delta^v f_{ij}) = \frac{\alpha_{ij}^h |\Delta^v f_{ij}|}{\alpha_{ij}^h + |\Delta^v f_{ij}|}. \tag{4.23}$$

This functional form (not including the coupling term) has been suggested and used by Geman and Reynolds in [14] for reasons similar to the ones detailed above.

The transition from ϕ^{GM} to ϕ^{GR} is shown by Rangarajan and Chellappa [45]. Consider the following modification of the prior

$$(\Delta^v f_{ij})^2(1 - l_{ij}^h)^2 + \alpha_{ij}^h(l_{ij}^h)^2 \quad \Rightarrow \quad |\Delta^v f_{ij}|(1 - l_{ij}^h)^2 + \alpha_{ij}^h(l_{ij}^h)^2,$$

with the line process l_{ij}^h still assuming continuous values in the unit interval $[0,1]$. The only change from the previous prior is the use of the absolute value of the

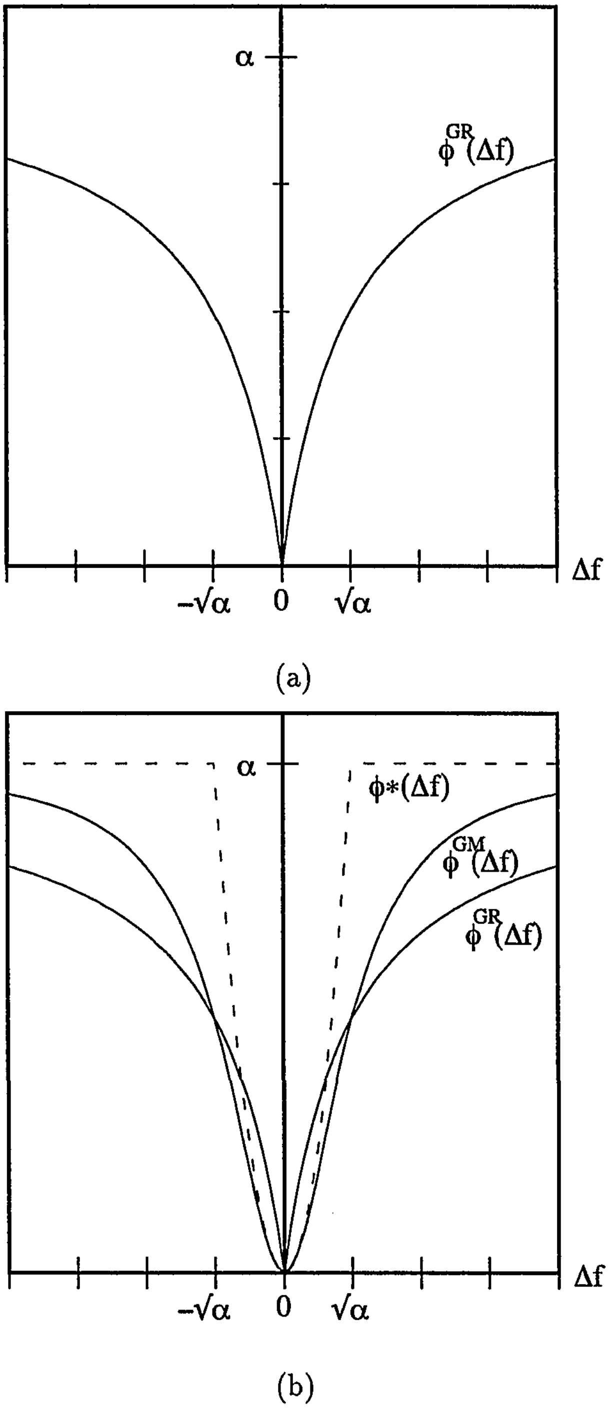


Figure 4.7: (a) ϕ^{GR}. (b) ϕ^*, ϕ^{GM}, and ϕ^{GR}.

spatial derivative instead of the squared form. The line process can be eliminated from this modified objective function in a manner similar to the earlier eliminations. When this is done and the solution substituted back into the objective function, we retrieve Equation 4.23. Thus the modification from ϕ^{GM} to ϕ^{GR} is not totally *ad hoc* but corresponds to a modification of the prior from the squared gradient form to the absolute value form. This is the prior used in our final objective function for this chapter. Figure 4.7(b) shows the relation between ϕ^{*}, ϕ^{GM}, and ϕ^{GR}.

After including the likelihood and our prior, we can arrive at the final objective function that we optimize:

$$\begin{aligned}
E_1^{GR}(\mathbf{f}) &= D(\mathbf{f}) + \lambda_1 \sum_{ij} \left[\phi^{GR}(\Delta^v f_{ij}) + \phi^{GR}(\Delta^h f_{ij})\right] \\
&= \sum_{t\theta} \left[\sum_{ij} \mathcal{H}_{t\theta;ij} f_{ij} - g_{t\theta} \log(\sum_{ij} \mathcal{H}_{t\theta;ij} f_{ij})\right] \\
&\quad + \lambda_1 \sum_{ij} \left[\frac{\left(\kappa_1(1-e^h_{ij}) + \kappa_2 e^h_{ij}\right) |\Delta^v f_{ij}|}{\left(\kappa_1(1-e^h_{ij}) + \kappa_2 e^h_{ij}\right) + |\Delta^v f_{ij}|}\right] \\
&\quad + \lambda_1 \sum_{ij} \left[\frac{\left(\kappa_1(1-e^v_{ij}) + \kappa_2 e^v_{ij}\right) |\Delta^h f_{ij}|}{\left(\kappa_1(1-e^v_{ij}) + \kappa_2 e^v_{ij}\right) + |\Delta^h f_{ij}|}\right].
\end{aligned} \qquad (4.24)$$

Note that the weighting factor λ_1 appears in the final objective function.

The objective function highlights two aspects of our approach. The first is the use of an intermodality term that modulates the creation of discontinuities according to *anatomical* as well as functional information. The second is the use of a cusp-like potential which alleviates the problems of non-convexity and hot spots.

We note that it is common in image processing and computer vision problems to include line process clique potentials to enforce higher level constraints that encourage discontinuities to occur along smooth, unbroken contours. Furthermore, the clique potentials discourage the formation of discontinuities along parallel, closely-spaced contours ("thick" contours). In particular, previous efforts [30] and [7] in anatomical–functional correlation have made use of clique potentials. Our results, however, show that the anatomical edge information $\mathbf{e}$ serves the above purposes.

In our formulation, the formation of smooth, unbroken contours is encouraged by the anatomical information $\mathbf{e}$ through the lowering of the penalty for creating a discontinuity. High values of $\mathbf{e}$ occur on a band of smooth contours and discontinuities tend to follow the locus of $\mathbf{e}$. However, the intrinsic properties of the weak membrane promote the formation of smooth, *thin* contours (see for example, in [5] on pages 87-88) rather than a *band* of contours. These intrinsic properties, "hysteresis" (formation of smooth and unbroken contours) and "non-maximum suppression" (formation of thin contours), are discussed extensively in [5].

Note also that in the absence of cliques, the line processes are independent and can thus be eliminated resulting in an equivalent energy function that depends only on the source intensities. This simplifies the search since we do not have to minimize the energy function with respect to the line processes.

For an actual minimization algorithm, the cusp-like potential used here is not differentiable and, hence, gradient–based minimization methods can not be used. Instead, we use a *quenching* algorithm which is identical to the zero temperature limit of simulated annealing. The details are presented in Section 4.4. It should be noted that the focus of the work in this chapter is on the incremental improvement in reconstruction due to the intermodality prior and that the issue of efficiency in the actual optimization algorithm is not considered.

4.3 Related Work

Chen, *et al.* [7] have incorporated CT/MR boundary information in PET reconstruction, using a Gibbs posterior distribution model based on the work of Johnson, *et al.* [21]. They approximated the Poisson likelihood with a Gaussian distribution, using the square root of the intensity as variables to be incorporated in a Gibbs posterior distribution model, and modeled the prior as a Gibbs distribution with line processes and cliques. Because of the difficulty of the mixed variable optimization, they let the line processes take continuous values. (There is more discussion on the work of Johnson, *et al.* [21] in Section 5.5.) An iterative conditional average (ICA) algorithm was used in which each intensity and line process value was updated by computing the mean values of the respective conditional distributions given the current estimates at each iteration. Intermodality information was incorporated by including a weighting factor in the conditional density of line sites. An edge value obtained from a registered CT/MR image and the value of the line process determined the weighting factor at each site. At each line site, if both the value of the line process (conditional mean value) and the intermodality edge value δ_{ij}

exceeded some threshold value, then the weighting factor was given by

$$w_{ij} = \left(\frac{l_{ij}\delta_{ij}}{l_{\mathrm{threshold}}\delta_{\mathrm{threshold}}} \right)^2 .$$

If either fell below its threshold value, the weighting factor was set to 1 and the conditional distribution of that line site was not influenced by the corresponding edge value of the anatomical image. The known edges of their phantom were used in the simulations as the intermodality information. Their algorithm is run some number of iterations without the intermodality edge information to estimate initial values of the line processes in the functional image. Although the results using the intermodality information show an improvement in the reconstruction, there is no mathematical justification corresponding to the modification of the conditional probability of the line process by including the modulated weighting factor. Furthermore, the line processes are explicitly included and the evaluation of the conditional probability of the line process makes the optimization step difficult. Since ICA is used, the mean values of the line processes may not necessarily approach 0 or 1.

Leahy and Yan [30] have also used the anatomical image information to improve PET reconstruction. They modeled the likelihood function as a Poisson distribution and the prior as a Gibbs distribution with line processes and simple cliques (simple hysteresis and non-maximum suppression). The anatomical edges were obtained by using an edge detector on a registered MR image. Using our energy function notation, their prior energy function can be rewritten

$$\begin{aligned} E_P &= \sum_{ij} \Big[(\Delta^v f_{ij})^2 (1 - l^h_{ij}) + \alpha l^h_{ij} + \alpha V_h(l^h_{ij}) \Big] + \\ &\quad \sum_{ij} \Big[(\Delta^h f_{ij})^2 (1 - l^v_{ij}) + \alpha l^v_{ij} + \alpha V_v(l^v_{ij}) \Big], \end{aligned}$$

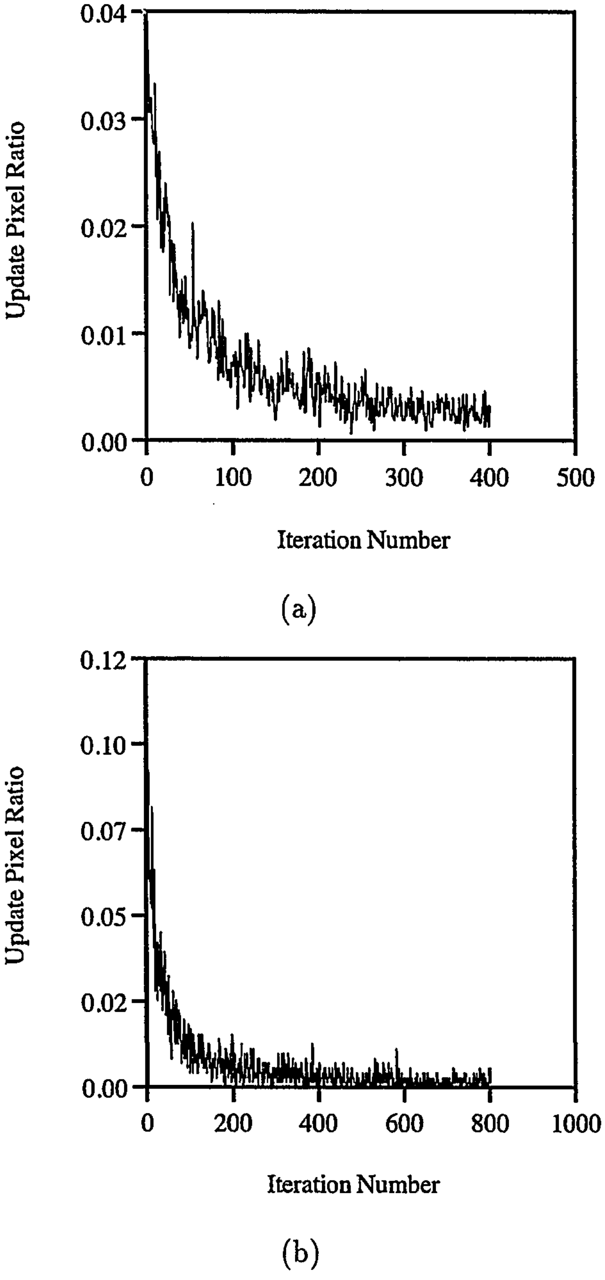
where the line clique potentials are given by:

$$\begin{aligned} V_h(l^h_{ij}) &= -\epsilon(l^h_{i,j+1} - l^h_{i+1,j}) l^h_{ij} = -\epsilon l^h_{i,j+1} l^h_{ij} + \epsilon l^h_{i+1,j} l^h_{ij}, \\ V_v(l^v_{ij}) &= -\epsilon(l^v_{i+1,j} - l^v_{i,j+1}) l^v_{ij} = -\epsilon l^v_{i+1,j} l^v_{ij} + \epsilon l^v_{i,j+1} l^v_{ij}. \end{aligned}$$

The first term in both line clique potentials corresponded to hysteresis, and the second term to non-maximum suppression. The anatomical edges were incorporated by treating the anatomical edges as *fixed* line processes in the functional image to be reconstructed. At the line sites where there existed corresponding anatomical edges, the line process values were fixed to the values of anatomical edges throughout the updating procedure. Only the line processes at sites without corresponding anatomical edges were considered for updates. The update was done using a modified ICA method. The modification of ICA algorithm involved introducing a temperature parameter in the conditional probability distribution of the line processes. They showed that this modification made the ICA method converge to a local minimum. (See also Section 5.5.) Taking anatomical edges and fixing them as function image line processes could, however, create false edges in the functional image due to the problems discussed in Section 4.1. There was no mathematical justification for their method of incorporating the intermodality information.

4.4 Simulation Results

The optimization was carried out via a quenching procedure. In our implementation, individual pixels are visited sequentially following a raster scan. One iteration in our algorithm corresponds to a single raster scan. At each pixel, the source intensity is altered and the change in objective is calculated. A move that reduces objective is always accepted, and a move that increases objective is always rejected. Pixels are altered according to the following scheme: a new value is sampled from a Gaussian probability density whose mean is equal to the current pixel intensity and whose standard deviation is fixed at 55. This analog value is added to the current pixel intensity and the result discretized to the nearest integer. The parameter space for **F** is thus discretized to 256 levels. Note that the allowed intensity range varies from 0 to 255 and when a sampled pixel value falls out of this range, we use a wraparound scheme wherein 255 is added to values less than 0 and subtracted from values greater than 255. The algorithm is run until the number of pixels whose values change in an iteration reaches a plateau of less than a suitably chosen threshold. By plateau we mean that the number of pixel changes per iteration is less than or equal to the threshold for 100 consecutive iterations. Figures 4.8(a) and (b) show the ratio of pixel acceptances per iteration in the quenching runs for reconstruction of Figure 4.9(g) and Figure 4.11(d), respectively.



(a)

(b)

Figure 4.8: Pixel acceptance ratio per iteration in the quenching run. (a) Ratio for Figure 4.9(g). (b) Ratio for Figure 4.11(d).

Following [16], we form an initial estimate using an Expectation Maximization–Maximum Likelihood (EM–ML) algorithm. In general, the EM–ML estimate is expected to get us close to the answer, thus saving an excessive number of iterations.

Phantom I

The first set of results is shown in Figure 4.9. Figure 4.9(a) shows a 64x64 brain phantom that was obtained from hand tracing selected anatomical edges of an actual MRI image of a human brain. The phantom has three regions that represent grey matter, white matter and a generalized internal structure. The outer region (grey matter) was completely and homogeneously filled with an equivalent radioisotope density of concentration 100; the intermediate region (white matter) was similarly filled with values of 25; and the two inner regions with values of 50. These regions crudely approximate the expected radioisotope density of a brain perfusion agent. Projection data was mathematically generated by calculating the Radon transform by using 64 equiangular projections around 360 degrees with 64 detector bins per projection. Independent Poisson noise was added to each detector bin. The total counts were approximately 3.3 x 10^6 counts.

Figure 4.9(b) shows the result obtained using filtered backprojection (FBP) with a ramp filter with no lowpass cutoff. Figure 4.9(c) shows the result obtained using EM–ML after 90 iterations where the total RMS error was least.

The next result, Figure 4.9(d) shows a weak membrane reconstruction. This is essentially an implementation of our reconstruction algorithm without benefit of the anatomical information and is obtained simply by setting $\dot{e}_{ij} = 0$ everywhere

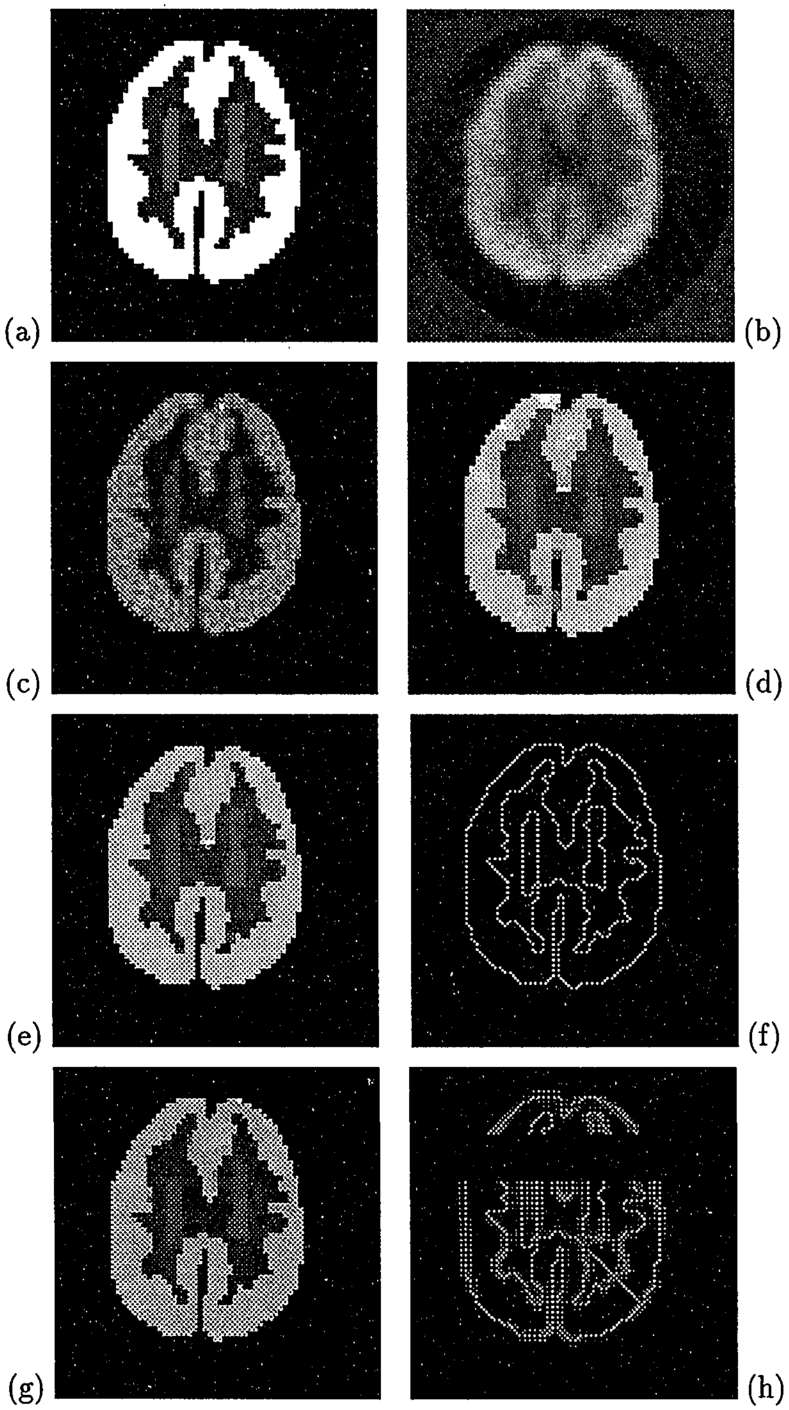


Figure 4.9: Simulations on Phantom I. (a) 64x64 brain phantom. (b) FBP reconstruction (c) EM–ML reconstruction. (d) Weak Membrane reconstruction. (e) Reconstruction using edge map (f). (f) Perfect edge map. (g) Reconstruction using edge map (h). (h) Blurred edge map with extra and missing edges.

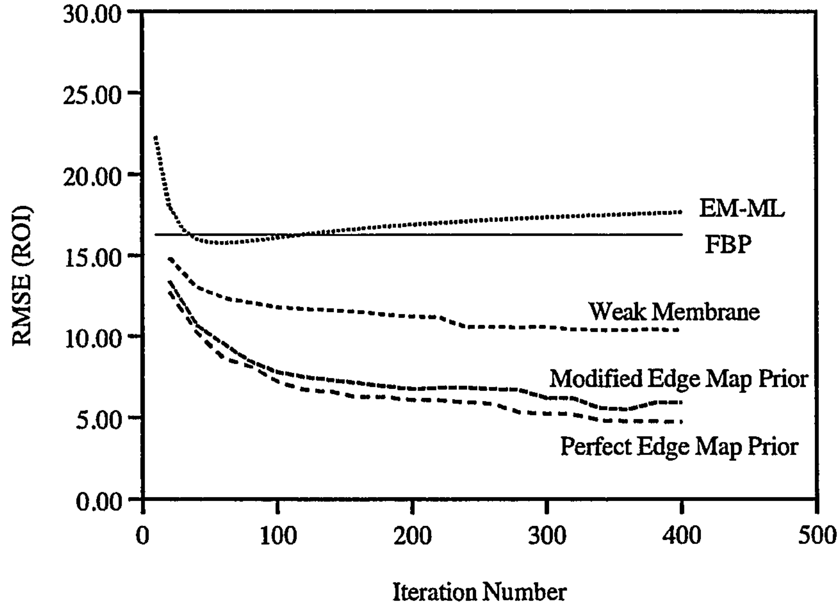
in our objective function (Equation 4.24). Values of $\kappa_1 = 13.0$ and $\lambda_1 = 0.69$ were used. The objective function no longer depends on κ_2 once e_{ij} is set to zero. The initial estimate was obtained with a choice of 25 iterations of the EM–ML algorithm. The weak membrane reconstruction was obtained using our quenching algorithm after 369 iterations. The plateau threshold of the quenching algorithm was 14.

The next result, Figure 4.9(e), demonstrates the improvement in the reconstruction due to the anatomical information. Here, we used our objective function (Equation 4.24) with $\kappa_1 = 15.0$ and $\kappa_2 = 5.0$, and $\lambda_1 = 0.69$. The anatomical information, illustrated in Figure 4.9(f), displays the combined vertical and horizontal edge maps. The value of e_{ij} was set to unity along the depicted boundaries and zero everywhere else. Again, the initial estimate was obtained with 25 iterations of the EM–ML algorithm, and a plateau threshold of 10 for the quenching algorithm, resulting in a total of 400 iterations.

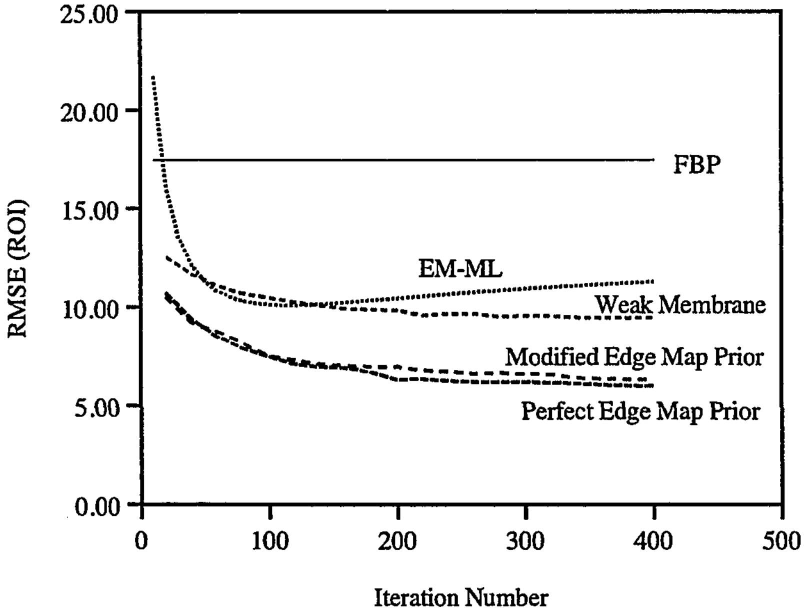
In order to be useful, the reconstruction algorithm should be robust with respect to uncertainties in the correlation between function and anatomy. The uncertainty can occur in several ways: (1) an anatomical edge is present but has no corresponding functional edge - extra edge; (2) a functional edge exists with no corresponding anatomical edge - missing edge; and (3) there is a misregistration error between the two types of edges. Note that extra edges can arise in two ways: either an anatomical edge is erroneously drawn, or the anatomical edge is correct but there is simply no corresponding functional edge at that location. Also, a missing edge can occur in two ways: either a significant anatomical edge is missed, or a strong functional edge occurs at an unexpected location.

An attempt to test the robustness of the algorithm with respect to these uncertainties is depicted in the modified edge maps of Figure 4.9(h). Here, there is an extra (diagonal) edge in the lower right corner, and an obvious band of missing edge information in the upper region. In addition, a general blurring of the edge map reflects our uncertainty in registration between the modalities. (For a real SPECT study, this blurring would translate into a misregistration error of about ±3mm.) A discussion of our model of this uncertainty is found in Section 4.1. In particular, $e^v_{ij} = 1$ at the same locations as in Figure 4.9(f) and for each such non-zero pixel, the horizontally, adjacent pixels $e^v_{i,j+1}$ and $e^v_{i,j-1}$ are set to 0.5. Similarly, $e^h_{ij} = 1$ at the same locations as in Figure 4.9(f) with vertically, adjacent pixels $e^h_{i+1,j}$ and $e^h_{i-1,j}$ set to 0.5. The corresponding reconstruction using this incomplete information is shown in Figure 4.9(g) with initial estimates as in Figure 4.9(e). The parameters κ_1, κ_2 and λ_1 for this case are the same for the reconstructions displayed in Figure 4.9(e). The same plateau threshold value of 10 was used and the total number of iterations was 409.

We now discuss the subjective quality of our results; later we present a quantitative assessment. Our best result using anatomical information, Figure 4.9(e), shows good image quality even in regions of low contrast. Detailed spatial structures appear well defined. The effect of a large band of missing edges as well as an extra diagonal line (lower right) is seen in Figure 4.9(g). Overall, the image quality in the region corresponding to the band of missing edges degrades slightly with some detail lost in the low contrast structures. There was no visible degradation



(a)



(b)

Figure 4.10: RMS error in ROI for Phantom I. (a) ROI corresponding to the outer bright region (grey matter). (b) ROI corresponding to the intermediate region (white matter).

due to the extra diagonal line. In addition, there is a slight loss of detail due to the blurring of the anatomical edges. Comparatively speaking, the weak membrane results in Figure 4.9(d) are more patchy, especially in regions where the phantom was uniform (e.g. grey matter). Small details, such as the border between the grey and white matter of the brain are not as well preserved in the weak membrane reconstruction. Not surprisingly, both the FBP and EM–ML reconstructions (Figures 4.9(b) and (c)) are noisier than the MAP reconstructions. Compared to the best intermodality reconstructions, the small, detailed structures in FBP and EM–ML are not as clearly resolved.

The RMS error relative to the phantom was evaluated for selected regions of interest (ROI) in order to quantitatively assess the efficacy of our algorithm. Figure 4.10(a) displays RMS error vs. iteration number for a ROI corresponding exactly to the bright, outer region (grey matter, value equal to 100) of the phantom. As seen, the Bayesian reconstructions result in improved RMS error in this region; the addition of prior anatomical information results in further improvement relative to the Bayesian reconstruction obtained without the anatomical information (weak membrane). Note also that the RMS error obtained with the modified (with blurred, missing and extra edges) edge map is nearly identical to the result obtained with the perfect edge map. This illustrates a robustness of the algorithm to slight errors in the edge map. Figure 4.10(b) shows a similar plot for the ROI corresponding to the intermediate (white matter, value equal to 25) region. Once again, the Bayesian reconstructions show considerable improvement over FBP and EM–ML, and the addition of anatomical prior information is clearly beneficial.

Phantom II

We also ran the algorithm on a second phantom shown in Figure 4.11(a). This 32 x 32 phantom consists of three regions of values 75, 130 and 150 against a background of 100. Projection data was mathematically generated by calculating the Radon transform using 32 equiangular projections around 180 degrees with 32 detector bins per projection. Independent Poisson noise was added to each detector bin. The total counts were approximately 1.3 x 10^6 counts.

Figure 4.11(b) shows the result obtained using EM–ML after 50 iterations, where the number of iterations was selected to correspond to the minimum RMS error relative to the phantom.

As with the previous phantom, we ran the Bayesian reconstructions with weak membrane and anatomical priors. For the anatomical priors, we used an edge map with missing and extra lines but without blur. Figure 4.11(e) show the anatomical edge map used in the intermodality reconstruction. Comparison with the phantom in Figure 4.11(a) shows a missing edge (upper right) and false edges corresponding to an extra region (center). The parameters for the weak membrane reconstruction were $\lambda_1 = 0.77$ and $\kappa_1 = 17.0$ with e_{ij} set to zero everywhere. Note that the objective function no longer depends on κ_2 once e_{ij} is set to zero. The parameters for the intermodality reconstruction were $\lambda_1 = 0.7$, $\kappa_1 = 15.0$ and $\kappa_2 = 5.0$.

Figure 4.11(c) shows the weak membrane reconstruction after 617 iterations. The reconstruction is generally smoother than FBP or EM–ML, but note the error

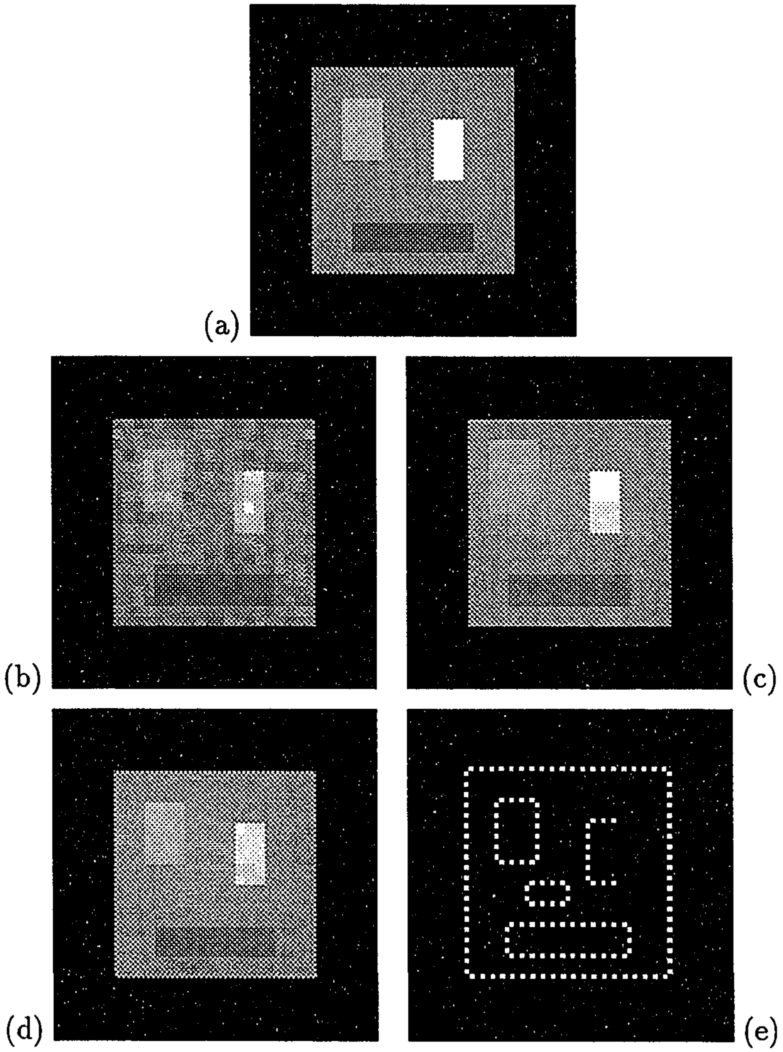


Figure 4.11: Simulation on Phantom II. (a) 32x32 mathematical phantom. (b) EM–ML reconstruction. (c) Weak Membrane reconstruction. (d) Reconstruction using edge map (e). (e) Edge map with an extra region and a missing edge.

in the reconstruction of the boundary of the upper left region. When anatomical information is employed as shown in Figure 4.11(d), the error is corrected and the resulting reconstruction is qualitatively superior. Figure 4.11(d) shows the result after 682 iterations. The plateau thresholds for both reconstructions were equal to 4.

As before, in order to quantitatively assess the performance of our algorithm, we plot the RMS error vs. iteration for two ROIs. Figure 4.12(a) displays RMS error vs. iteration number for an ROI corresponding exactly to the bright upper right region of the phantom; Figure 4.12(b) shows the RMS error for an ROI corresponding exactly to the background (i.e. the region in the phantom with value 100). As seen, the Bayesian reconstructions result in improvement in these regions; the addition of prior anatomical information results in further improvement relative to the reconstruction obtained without the anatomical information (weak membrane).

Phantom III

In proceeding to a more realistic case, we used a Monte Carlo simulation program [61], [60], [59] that simulates SPECT projection data from a three dimensional phantom. It simulates the physical absorption and scattering properties of tissue, as well as geometric imaging factors and detector response. The gray levels in Figure 4.13(a) depict the relative radionuclide concentrations (ratio 6:4:1 for inner structure, grey matter and white matter, respectively) used in the simulation. To realistically model scattering, the source distribution was made three dimensional by stacking the two dimensional source Figure 4.13(a) in the axial direction. The

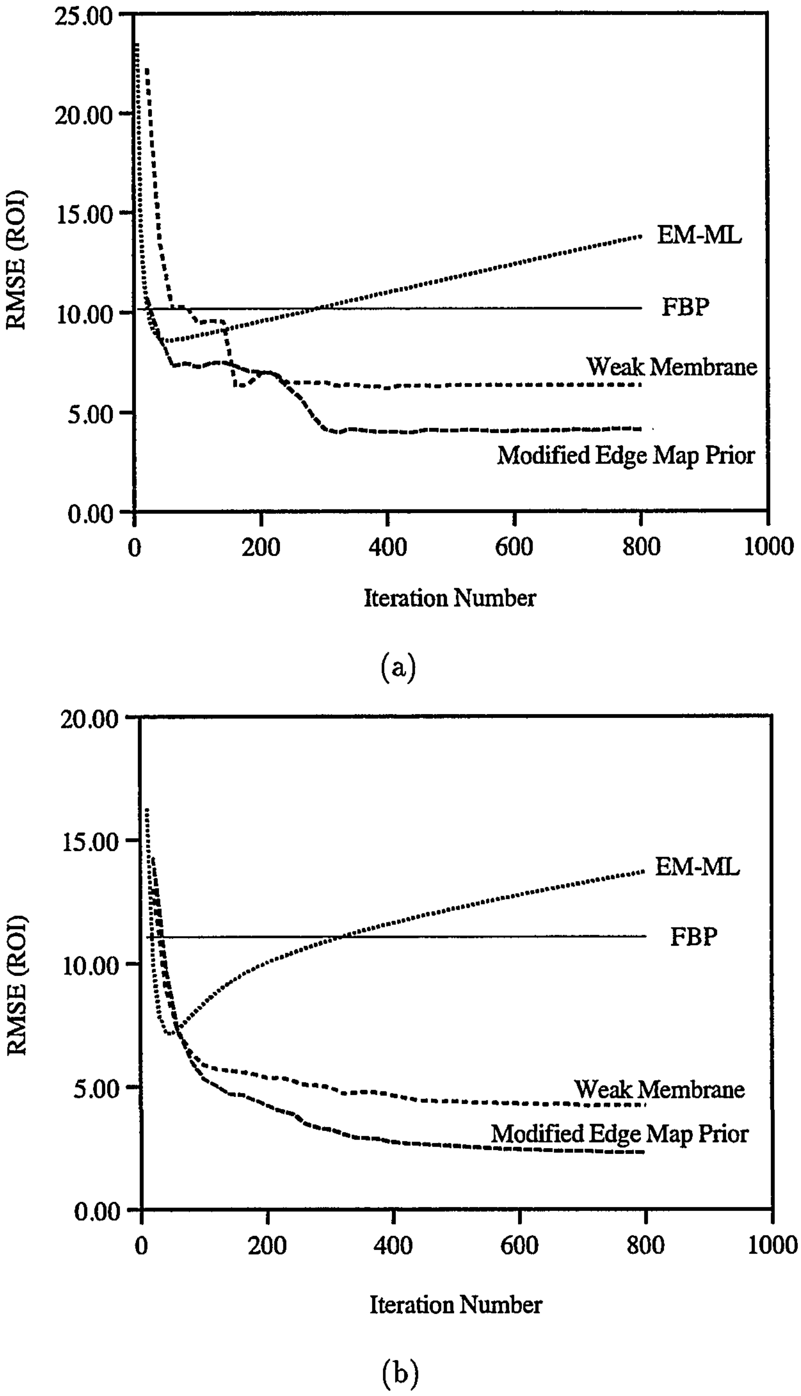


Figure 4.12: RMSE error in ROI for phantom II. (a) ROI corresponding to the bright upper right region. (b) ROI corresponding to the background region.

primary emission energy was set to 140 Kev which corresponds to 99mTc. The parameters of the Monte Carlo program were set so that the voxel size was 1.5 x 1.5 x 7.5 mm on a side. Within the border of the phantom, water was assumed to be the uniform attenuating medium. The collimator was modeled to correspond to a Low Energy All Purpose (LEAP) collimator with an acceptance angle of 1.5 degrees FWHM. Given the source characteristics and the geometry, the typical SNR of the detector counts was approximately 7.0.

The resulting reconstructions are shown in Figure 4.13. Figure 4.13(a) shows the phantom source distribution. Figure 4.13(b) shows a conventional filtered backprojection reconstruction using a ramp filter with no lowpass cutoff. Figure 4.13(c) shows an EM–ML reconstruction after 50 iterations. Figure 4.13(d) shows our reconstruction using intermodality information. The intermodality information was the same as in Figure 4.9(g). The parameters for our reconstruction were $\lambda_1 = 0.2$, $\kappa_1 = 100.0$, $\kappa_2 = 5.0$ and the initial estimate was obtained with 10 iterations of the EM–ML algorithm. For our quenching algorithm, the range of pixels sampled was 0 to 350 and a standard deviation of 10 was chosen. No attenuation or scatter correction was applied in any of the three algorithms above. Although the intermodality reconstruction is less noisy and exhibits somewhat improved recovery of spatial detail, it is more patchy. To improve the reconstruction, we must account for attenuation and scattering. This requires a more realistic forward transformation model.

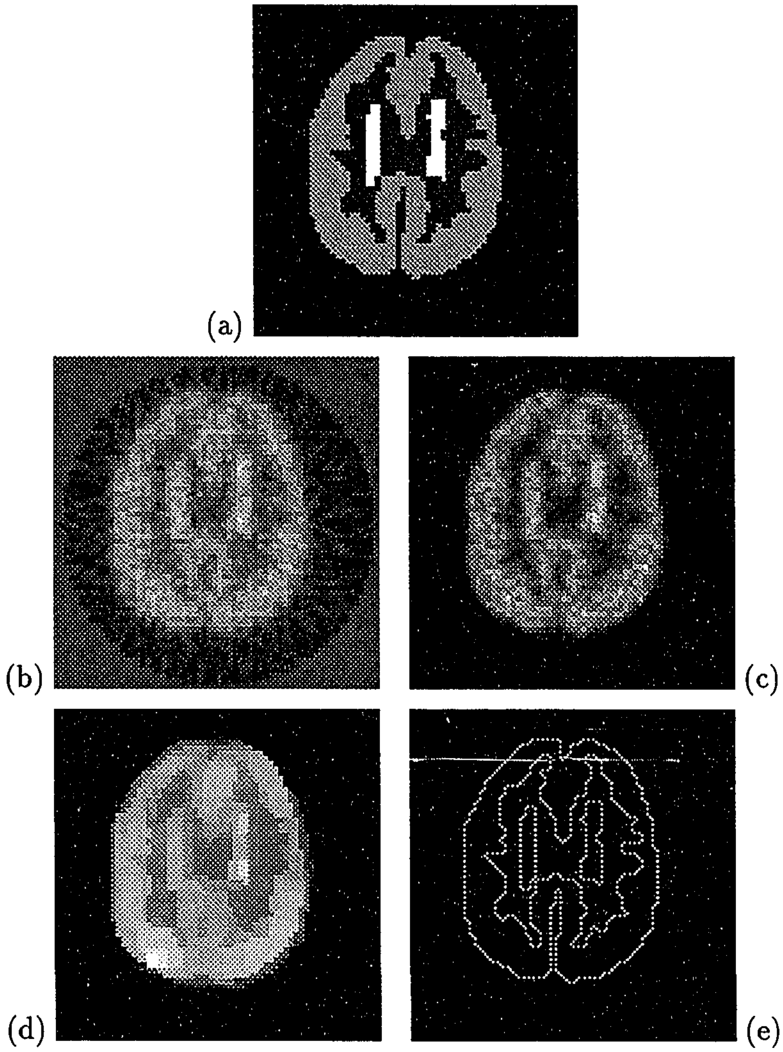


Figure 4.13: Simulation on Phantom III. (a) 64x64 brain phantom. (b) FBP reconstruction (c) EM–ML reconstruction. (d) Reconstruction using edge map (e). (e) Perfect edge map.

4.5 Discussion

We have presented in this chapter a Bayesian model that incorporates correlated anatomical edge information into the functional image reconstruction. Our results, obtained by using mathematical phantoms, show that the anatomical information leads to improved reconstructions. On the other hand, we need a more realistic forward operator, $\mathcal{H}$, with attenuation corrections to fully test the method on clinical data. The reconstruction method has shown some degree of robustness to errors such as blurring, and missing or extra edges.

An important feature of our approach is that anatomical information is incorporated independently of an actual optimization technique. This takes the form of a space varying modulation of the penalty for creating discontinuities through the use of a coupling term. A second advantage is in the use of a cusp-like potential, the properties of which were discussed in Section 4.2.

Several issues must be addressed before it is practical to use anatomical information as priors. Further work must be done to test validity of correlating anatomical structures with function for different radiopharmaceuticals. Given reliable knowledge of anatomy–function correlation, there must be efficient procedure to extract significant anatomical edges that are likely correlate to function. This will require the ability to reliably distinguish various tissue types in an anatomical image. In addition, accurate 3-D registration is required along with an improved means of modeling residual registration errors into the reconstruction procedure.

The work detailed in this chapter presents one possible means of incorporating anatomical information into a functional image reconstruction problem.

Chapter 5

A Continuation Method for Bayesian Reconstruction

As shown in Chapter 4, a Bayesian reconstruction is often modeled with a Gibbs prior defined on both the binary valued line processes, **l**, as well as the source intensities, **f**. Due to the mixed (continuous and binary) variables, however, the simultaneous MAP estimation of **f** and **l** becomes a difficult problem. In Chapter 4, we used a Gibbs "potential" function, ϕ^{GR} obtained by making some modifications to our original function, ϕ^*. The function ϕ^{GR} implicitly captures the property of line processes, but only in an approximate manner. The "correct" ϕ function, ϕ^*, consistent with the use of line processes, however, leads to difficult minimization problems.

The description in this chapter expands the work in [31].

In this chapter, we present a new framework for a Bayesian reconstruction in emission tomography that offers a principled and efficient way of handling the problems associated with the mixed variable objectives. Here, a continuation method is used to approach the "correct" potential function, ϕ^*, through a *sequence* of smooth ϕ functions – to this end, a deterministic annealing algorithm is derived and employed. This results in a *sequence* of objective functions (depending only on the continuous variables) that approaches the original mixed variable objective function. The sequence is indexed by a control parameter (the temperature). At each temperature, a standard descent optimization algorithm is used to find a solution which is then used as an initial estimate for the next temperature setting. The objective functions at high temperatures are smooth approximations of the objective functions at lower temperatures. Consequently, it is easier to minimize the objective functions at high temperatures and then track the minimum through variation of the temperature. This is the essence of a *continuation* method. In other domains, such as computer vision, where certain problems can be formulated employing Bayesian estimates, continuation methods have been successfully used [13, 45] to find good solutions.

In Chapter 4, we made use of intermodality information to model the prior distribution. For the sake of simplicity in developing the continuation method, we only use the weak membrane model without intermodality information. It should be noted, however, that intermodality information can be used in the continuation framework that we develop here.

Table 5.1: Symbols used in Chapter 5

λ_2	Prior weighting parameter
α	Space invariant weak membrane parameter
β	Temperature control parameter
$\mathbf{z}^h, \mathbf{z}^v$	Expected value of line processes
$E_{P2}(\mathbf{f}, \mathbf{l})$	Energy function for the prior on **f** and **l**
$E_{MP}(\mathbf{f})$	Energy function for the marginal intensity prior
$E_{MP}(\mathbf{f}, \beta)$	Energy function for the temperature modified marginal intensity prior
$E_2(\mathbf{f}, \mathbf{l})$	Overall objective function
$E_2^*(\mathbf{f})$	Equivalent overall objective function obtained by eliminating line process
$E_\beta(\mathbf{f})$	Temperature modified objective function used in optimization at each β
$\phi^*(\cdot)$	Prior potential function obtained by eliminating line process; broken parabola
$\phi_{MP}(\cdot)$	Marginal intensity prior potential function
$\phi_\beta(\cdot)$	Temperature modified prior potential function

5.1 An Energy Function for MAP Estimation

We are interested in a MAP estimate of a Bayesian model for the reconstruction problem. As in Section 4.2.2, we have the following for the MAP estimate

$$\begin{aligned}(\hat{\mathbf{f}}_{\mathrm{MAP}}, \hat{\mathbf{l}}_{\mathrm{MAP}}) &= \arg\max_{(\mathbf{f},\mathbf{l})} \Pr(\mathbf{F}=\mathbf{f},\mathbf{L}=\mathbf{l}|\mathbf{G}=\mathbf{g}) \\ &= \arg\min_{(\mathbf{f},\mathbf{l})} \left[-\log\Pr(\mathbf{G}=\mathbf{g}|\mathbf{F}=\mathbf{f},\mathbf{L}=\mathbf{l}) - \log\Pr(\mathbf{F}=\mathbf{f},\mathbf{L}=\mathbf{l})\right] \\ &= \arg\min_{(\mathbf{f},\mathbf{l})} \left[-\log\Pr(\mathbf{G}=\mathbf{g}|\mathbf{F}=\mathbf{f}) - \log\Pr(\mathbf{F}=\mathbf{f},\mathbf{L}=\mathbf{l})\right] \qquad (5.1)\end{aligned}$$

As in Chapter 4, the likelihood term $\Pr(\mathbf{G}=\mathbf{g}|\mathbf{F}=\mathbf{f},\mathbf{L}=\mathbf{l})$ becomes $\Pr(\mathbf{G}=\mathbf{g}|\mathbf{F}=\mathbf{f})$ since the projection data, $\mathbf{g}$, obtained from a forward projection of the source, $\mathbf{f}$, does not involve the line process, $\mathbf{l}$.

The likelihood is again modeled as a Poisson distribution as in Equation 3.8. From Equation 4.6, we have

$$\begin{aligned}&-\log\Pr(\mathbf{G}=\mathbf{g}|\mathbf{F}=\mathbf{f}) \\ &\quad= D(\mathbf{f}) \;+\; \text{terms independent of } \mathbf{f} \\ &\quad= \sum_{t\theta}\sum_{ij}\mathcal{H}_{t\theta;ij}f_{ij} - \sum_{t\theta} g_{t\theta}\log(\sum_{ij}\mathcal{H}_{t\theta;ij}f_{ij}) \;+\; \text{terms independent of } \mathbf{f}.\end{aligned}$$

We model the prior as a Gibbs function defined over the intensities and the line processes:

$$\Pr(\mathbf{F}=\mathbf{f},\mathbf{L}=\mathbf{l}) = \frac{1}{Z_4}e^{-E_{P2}(\mathbf{f},\mathbf{l})} \qquad (5.2)$$

where the partition function, Z_4, is a constant independent of $\mathbf{f}$ and $\mathbf{l}$. The source distribution is assumed to be piecewise smooth. We again use the *weak membrane* model in Equation 4.12 for the prior. The associated energy function for our prior can then be written

$$E_{P2}(\mathbf{f},\mathbf{l}) \stackrel{\text{def}}{=} \lambda_2 \sum_{ij} \left[\left((\Delta^v f_{ij})^2 (1 - l^h_{ij}) + \alpha l^h_{ij} \right) + \left((\Delta^h f_{ij})^2 (1 - l^v_{ij}) + \alpha l^v_{ij} \right) \right]. \quad (5.3)$$

The two terms involving $\Delta^v f$ and $\Delta^h f$ encourage smoothness except where discontinuities occur. The terms with the parameter α penalize the creation of discontinuities. Fewer discontinuities are obtained at large values of α. Note that here the prior energy function, $E_{P2}(\mathbf{f},\mathbf{l})$, contains the weighting parameter λ_2. It is modeled somewhat differently than in Equation 4.9.

We can now write the MAP estimate

$$\begin{aligned}
(\hat{\mathbf{f}}_{\text{MAP}}, \hat{\mathbf{l}}_{\text{MAP}}) &= \arg\min_{(\mathbf{f},\mathbf{l})} \left[-\log \Pr(\mathbf{G} = \mathbf{g} | \mathbf{F} = \mathbf{f}) - \log \Pr(\mathbf{F} = \mathbf{f}, \mathbf{L} = \mathbf{l}) \right] \\
&= \arg\min_{(\mathbf{f},\mathbf{l})} \left[D(\mathbf{f}) + E_{P2}(\mathbf{f},\mathbf{l}) \right] \\
&= \arg\min_{(\mathbf{f},\mathbf{l})} E_2(\mathbf{f},\mathbf{l}),
\end{aligned}$$

where the overall objective function

$$\begin{aligned}
E_2(\mathbf{f},\mathbf{l}) &\stackrel{\text{def}}{=} D(\mathbf{f}) + E_{P2}(\mathbf{f},\mathbf{l}) \\
&= \sum_{t\theta} \sum_{ij} \mathcal{H}_{t\theta;ij} f_{ij} - \sum_{t\theta} g_{t\theta} \log\Big(\sum_{ij} \mathcal{H}_{t\theta;ij} f_{ij}\Big)
\end{aligned}$$

$$+\lambda_2 \sum_{ij} \left[\left((\Delta^v f_{ij})^2 (1 - l^h_{ij}) + \alpha l^h_{ij} \right) + \left((\Delta^h f_{ij})^2 (1 - l^v_{ij}) + \alpha l^v_{ij} \right) \right] . \quad (5.4)$$

The parameter λ_2 weights the prior with respect to the likelihood objective $D(\mathbf{f})$.

5.2 A Continuation Method for MAP Estimation

In principle, one could minimize the non-convex energy function in Equation 5.4 by searching over all configurations of $(\mathbf{f}, \mathbf{l})$, but this is impractical due to the large number of configurations. The search over configurations in the weak membrane, however, can be reduced to a search only over the intensities, $\mathbf{f}$. This was shown in Section 4.2.4 for the prior distribution with space varying weak membrane parameters, α^h_{ij} and α^v_{ij}. Here, however, we are interested in the weak membrane prior without intermodality information. Blake and Zisserman [5] have shown for the weak membrane prior that the solution for the line process $\hat{l}^h_{ij}$, (the MAP estimate), reduces to

$$\hat{l}^h_{ij} = \begin{cases} 0 & (\Delta^v f_{ij})^2 \leq \alpha \\ 1 & (\Delta^v f_{ij})^2 > \alpha \end{cases} \quad (5.5)$$

with a similar expression for the vertical line process l^v_{ij}. Note that α is a space-invariant parameter. When the MAP estimate for $\mathbf{l}$ is substituted back into the objective function of Equation 5.4, we obtain an *equivalent* objective function that depends only on the intensities:

$$E_2^*(\mathbf{f}) = D(\mathbf{f}) + \sum_{ij} \left[\phi^*(\Delta^v f_{ij}) + \phi^*(\Delta^h f_{ij}) \right], \tag{5.6}$$

where

$$\phi^*(\Delta^v f_{ij}) = \begin{cases} \lambda_2 (\Delta^v f_{ij})^2 & (\Delta^v f_{ij})^2 \le \alpha \\ \lambda_2 \alpha & (\Delta^v f_{ij})^2 > \alpha \end{cases}, \tag{5.7}$$

with a corresponding expression for $\phi^*(\Delta^h f_{ij})$. This is similar to the case in Chapter 4, but we repeat it here for notational consistency. Figure 5.1 shows $\phi^*(\Delta f_{ij})$. We again refer to ϕ^* as a broken parabola as in Chapter 4. Though the global minima of Equation 5.6 are the same as those of Equation 5.4 and the use of binary line variables have been circumvented, there are problems associated with minimizing an objective function consisting of the likelihood and the broken parabola. The objective function remains non-convex (discussed in Section 4.2.4) and, since the broken parabola makes the objective function non-differentiable, efficient methods of minimization such as gradient descent are ruled out. In Chapter 4, we dealt with these problems by approximating ϕ^* with ϕ^{GR}. In this chapter we pursue a new approach – a continuation method – where a sequence of functions is used to closely approximate ϕ^*. In this new approach, we are trying to find a solution to the original problem and not simply to approximate it with modification steps.

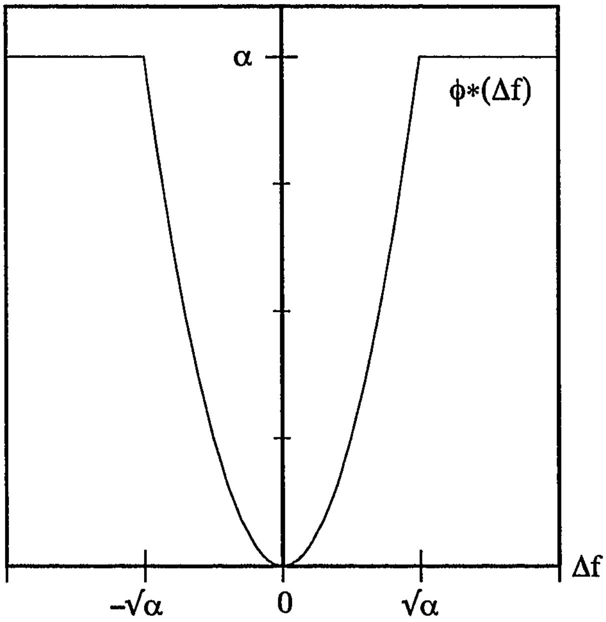


Figure 5.1: $\phi^*(\Delta f)$

5.2.1 Marginal Posterior and Marginal Prior

We first show the relationship between the broken parabola and the potential function arising out of the marginal prior. The marginal intensity posterior distribution can be written

$$
\begin{aligned}
\Pr(\mathbf{F}=\mathbf{f}|\mathbf{G}=\mathbf{g}) &= \sum_{\{\mathbf{l}\}} \Pr(\mathbf{F}=\mathbf{f},\mathbf{L}=\mathbf{l}|\mathbf{G}=\mathbf{g}) \\
&= \sum_{\{\mathbf{l}\}} \frac{\Pr(\mathbf{G}=\mathbf{g}|\mathbf{F}=\mathbf{f})\Pr(\mathbf{F}=\mathbf{f},\mathbf{L}=\mathbf{l})}{\Pr(\mathbf{G}=\mathbf{g})} \\
&= \frac{\Pr(\mathbf{G}=\mathbf{g}|\mathbf{F}=\mathbf{f})}{\Pr(\mathbf{G}=\mathbf{g})} \left[\sum_{\{\mathbf{l}\}} \Pr(\mathbf{F}=\mathbf{f},\mathbf{L}=\mathbf{l})\right]
\end{aligned}
$$

$$= \frac{\Pr(\mathbf{G}=\mathbf{g}|\mathbf{F}=\mathbf{f})}{\Pr(\mathbf{G}=\mathbf{g})}\Pr(\mathbf{F}=\mathbf{f}), \tag{5.8}$$

where $\Pr(\mathbf{F}=\mathbf{f})$ is the marginal intensity prior. We can evaluate the marginal intensity prior using the prior distribution in Equation 5.2 and integrating out the line processes:

$$\begin{aligned}
\Pr(\mathbf{F}=\mathbf{f}) &= \sum_{\{\mathbf{l}\}}\Pr(\mathbf{F}=\mathbf{f},\mathbf{L}=\mathbf{l}) && (5.9)\\
&= \sum_{\{l^v_{ij}\},\{l^h_{ij}\}}\frac{1}{Z_4}e^{-\lambda_2\sum_{ij}\left\{[(\Delta^v f_{ij})^2(1-l^h_{ij})+\alpha l^h_{ij}]+[(\Delta^h f_{ij})^2(1-l^v_{ij})+\alpha l^v_{ij}]\right\}} && (5.10)\\
&= \frac{1}{Z_4}\prod_{ij}\left[\sum_{l^h_{ij}=\{0,1\}}e^{-\lambda_2\{(\Delta^v f_{ij})^2(1-l^h_{ij})+\alpha l^h_{ij}\}}\right]\left[\sum_{l^v_{ij}=\{0,1\}}e^{-\lambda_2\{(\Delta^h f_{ij})^2(1-l^v_{ij})+\alpha l^v_{ij}\}}\right] && (5.11)\\
&= \frac{1}{Z_4}\prod_{ij}\left[e^{-\lambda_2(\Delta^v f_{ij})^2}+e^{-\lambda_2\alpha}\right]\left[e^{-\lambda_2(\Delta^h f_{ij})^2}+e^{-\lambda_2\alpha}\right] && (5.12)\\
&= \frac{1}{Z_4}e^{\sum_{ij}\left\{\log[e^{-\lambda_2(\Delta^v f_{ij})^2}+e^{-\lambda_2\alpha}]+\log[e^{-\lambda_2(\Delta^h f_{ij})^2}+e^{-\lambda_2\alpha}]\right\}} && (5.13)\\
&\stackrel{\text{def}}{=} \frac{1}{Z_4}e^{-\left\{\sum_{ij}\phi_{MP}(\Delta^v f_{ij})+\sum_{ij}\phi_{MP}(\Delta^h f_{ij})\right\}} && (5.14)\\
&\stackrel{\text{def}}{=} \frac{1}{Z_4}e^{-E_{MP}(\mathbf{f})}.
\end{aligned}$$

In Equation 5.10, $\sum_{\{l^v_{ij}\},\{l^h_{ij}\}}$ denotes the summation over all configurations of l^v and l^h, where l^v_{ij} and l^h_{ij} are either 0 or 1 for all (i,j). The transition from Equation 5.10 to Equation 5.11 can be shown by analogy with its continuous variables equivalent, i.e. $\int\cdot\cdot\int e^{-\sum x_i^2}dx_1\cdot\cdot dx_n = \prod_i\int e^{-x_i^2}dx_i$. Equation 5.12 is obtained by summing over l^h_{ij} and l^v_{ij} (taking values 0 and 1). Taking the logarithm of the product in Equation 5.12 and then exponentiating it, we obtain Equation 5.13. Shown in

Equation 5.14, the marginal intensity prior $\Pr(\mathbf{F} = \mathbf{f})$ is a Gibbs distribution with an energy function

$$E_{MP}(\mathbf{f}) \overset{\text{def}}{=} \sum_{ij} \phi_{MP}(\Delta^v f_{ij}) + \sum_{ij} \phi_{MP}(\Delta^h f_{ij}),$$

where the potential function

$$\begin{aligned} \phi_{MP}(\Delta^v f_{ij}) &\overset{\text{def}}{=} -\log\left[e^{-\lambda_2(\Delta^v f_{ij})^2} + e^{-\lambda_2\alpha}\right] \\ \phi_{MP}(\Delta^h f_{ij}) &\overset{\text{def}}{=} -\log\left[e^{-\lambda_2(\Delta^h f_{ij})^2} + e^{-\lambda_2\alpha}\right]. \end{aligned} \tag{5.15}$$

By examining $\phi_{MP}(\Delta f_{ij})$, we see that the above marginal prior distribution $\Pr(\mathbf{F} = \mathbf{f})$ is a smooth Gibbs prior. Figure 5.2 shows that the function, $\phi_{MP}(\Delta f_{ij})$, is indeed a smooth approximation of the broken parabola, $\phi^*(\Delta f_{ij})$, in Figure 5.1.

We can now write the MAP estimate for the marginal posterior

$$\begin{aligned} &\arg\max_{\mathbf{f}} \Pr(\mathbf{F} = \mathbf{f} | \mathbf{G} = \mathbf{g}) \\ &\quad = \arg\min_{\mathbf{f}} [-\log \Pr(\mathbf{F} = \mathbf{f} | \mathbf{G} = \mathbf{g})] \\ &\quad = \arg\min_{\mathbf{f}} [-\log \Pr(\mathbf{G} = \mathbf{g} | \mathbf{F} = \mathbf{f}) - \log \Pr(\mathbf{F} = \mathbf{f}) \\ &\qquad\qquad + \text{terms independent of } \mathbf{f}] \\ &\quad = \arg\min_{\mathbf{f}} [D(\mathbf{f}) + E_{MP}(\mathbf{f})] \\ &\quad = \arg\min_{\mathbf{f}} \left[D(\mathbf{f}) + \sum_{ij} \left[\phi_{MP}(\Delta^v f_{ij}) + \phi_{MP}(\Delta^h f_{ij})\right]\right]. \end{aligned}$$

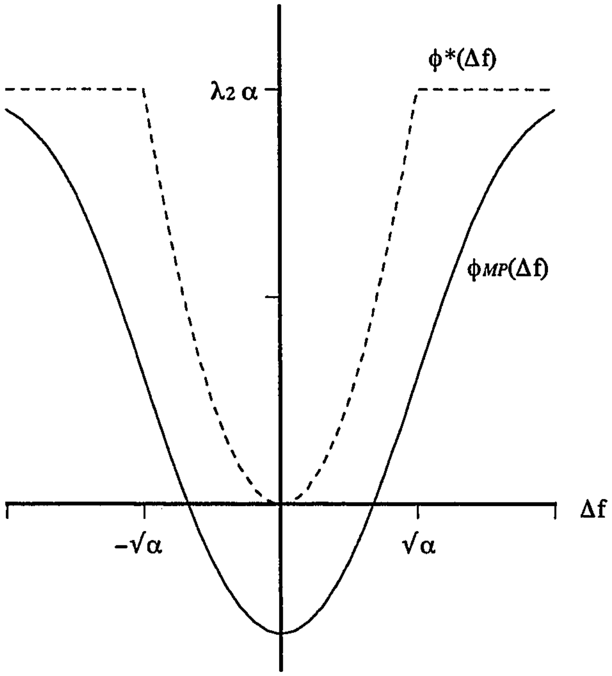


Figure 5.2: ϕ_{MP}

Note that the above objective function, $D(\mathbf{f}) + E_{MP}(\mathbf{f})$, approximates the original objective, $E_2^*(\mathbf{f})$, in Equation 5.6 with $\phi_{MP}(\Delta f_{ij})$, which is a smooth approximation of the broken parabola $\phi^*(\Delta f_{ij})$. This gives us the motivation to use a continuation method approach to our optimization problem.

Since ϕ_{MP} is a smooth, differentiable approximation of the broken parabola, the minimization of the objective function is easier since gradient descent methods can be used. Note, however, that solutions to the MAP estimate on marginal posterior are not the solutions to the original objective function. If we can generate a sequence of smooth and differentiable ϕ functions from ϕ_{MP} that more and more closely approach ϕ^*, then we expect the solution to this sequence of problems to tend to the solution of the original ϕ^* problem.

5.2.2 Introducing a Temperature Parameter

We can indeed generate a sequence of ϕ's from ϕ_{MP} that approach the broken parabola, ϕ^*. This is achieved by introducing a control parameter, β, usually identified with the inverse of a computational temperature ($\beta = \frac{1}{T}$), into the potential function, ϕ_{MP}, in Equation 5.15. This parameter gives the extra degree of freedom necessary to apply the continuation method. We will refer to β as a temperature control parameter and $T = \frac{1}{\beta}$ as a temperature. Consider the following sequence of smooth potential functions indexed by β:

$$\phi_\beta(\Delta f_{ij}) \stackrel{\text{def}}{=} -\frac{1}{\beta}\log\left[e^{-\beta\lambda_2(\Delta f_{ij})^2} + e^{-\beta\lambda_2\alpha}\right]. \tag{5.16}$$

Note that for $\beta = 1$, $\phi_{\beta=1}(\Delta f_{ij}) = \phi_{MP}(\Delta f_{ij})$. We now examine the relationship between the broken parabola $\phi^*(\Delta f_{ij})$, and the above potential function $\phi_\beta(\Delta f_{ij})$.

Consider the situation at large values of β (low temperature) for the potential function $\phi_\beta(\Delta f_{ij})$ in Equation 5.16. If $\alpha > (\Delta f_{ij})^2$, the first term, $e^{-\beta\lambda_2(\Delta f_{ij})^2}$, inside the log$(\cdot)$ expression dominates. This reduces the entire log expression to $\lambda_2(\Delta f_{ij})^2$. If $\alpha < (\Delta f_{ij})^2$, the second term, $e^{-\beta\lambda\alpha}$, inside the log$(\cdot)$ expression dominates and the entire expression reduces to $\lambda_2\alpha$. Inspection reveals that at large values of β, ϕ_β is in fact the broken parabola ϕ^* of Equation 5.7. At low values of β, the function is very smooth. As can be seen from Figure 5.3, ϕ_β

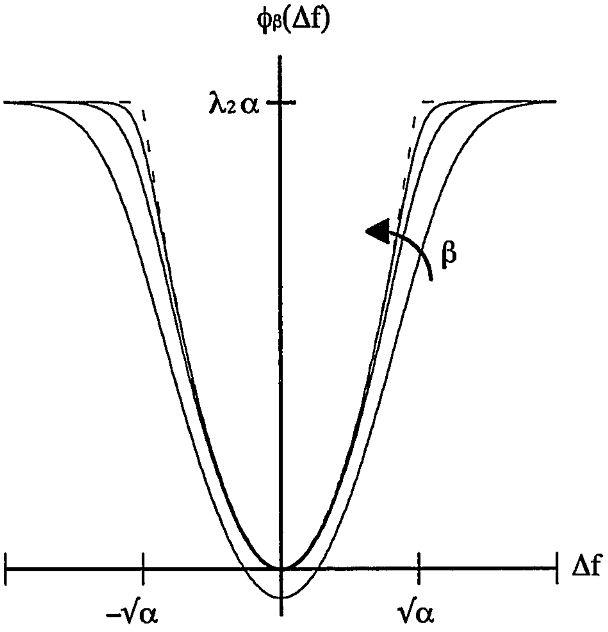


Figure 5.3: ϕ_β functions with $\beta = 2, 4, 8$

approaches the broken parabola as the temperature is reduced to zero (annealing).

Our goal is to find the minimum of the objective function

$$
\begin{aligned}
E_\beta(\mathbf{f}) &= D(\mathbf{f}) + E_{\mathrm{MP}}(\mathbf{f};\beta) \qquad (5.17)\\
&= D(\mathbf{f}) + \sum_{ij}\left[\phi_\beta(\Delta^v f_{ij}) + \phi_\beta(\Delta^h f_{ij})\right]\\
&= \sum_{t\theta}\sum_{ij}\mathcal{H}_{t\theta;ij}f_{ij} - \sum_{t\theta} g_{t\theta}\log(\sum_{ij}\mathcal{H}_{t\theta;ij}f_{ij})\\
&\quad + \left(\frac{-1}{\beta}\right)\sum_{ij}\left[\log\left(e^{-\beta\lambda_2(\Delta^v f_{ij})^2} + e^{-\beta\lambda_2\alpha}\right) + \log\left(e^{-\beta\lambda_2(\Delta^v f_{ij})^2} + e^{-\beta\lambda_2\alpha}\right)\right]
\end{aligned}
$$

by performing a standard minimization procedure at each temperature. This minimum is then used as an initial condition for the next temperature setting. Because the potential function ϕ_β is smooth and differentiable, we are able to employ gradient descent-like methods for the minimization.

The temperature modified distribution

We would like to show now that $\phi_\beta(\Delta f_{ij})$ in Equation 5.16 can be obtained from the temperature modified prior distribution. Consider the following modification [13] of the prior $\Pr(\mathbf{F} = \mathbf{f}, \mathbf{L} = \mathbf{l})$ in Equation 5.2 where we introduce the temperature parameter ($\beta > 0$) according to :

$$\begin{aligned} \Pr_\beta(\mathbf{F} = \mathbf{f}, \mathbf{L} = \mathbf{l}) &\stackrel{\text{def}}{=} \frac{1}{Z_5}[\Pr(\mathbf{F} = \mathbf{f}, \mathbf{L} = \mathbf{l})]^\beta \\ &= \frac{1}{Z_6} e^{-\beta E_{P2}(\mathbf{f},\mathbf{l})}. \end{aligned} \tag{5.18}$$

Here, Z_5 is a normalizing constant and Z_6 is a new partition function, both independent of $\mathbf{f}$ and $\mathbf{l}$. For $\beta = 1$, $\Pr_{\beta=1}(\mathbf{F} = \mathbf{f}, \mathbf{L} = \mathbf{l}) = \Pr(\mathbf{F} = \mathbf{f}, \mathbf{L} = \mathbf{l})$. At high values of β, the modified prior becomes more peaked. The locations of the peaks and valleys, however, are independent of the value of β. Therefore, both $\Pr_\beta(\mathbf{F} = \mathbf{f}, \mathbf{L} = \mathbf{l})$ and $\Pr(\mathbf{F} = \mathbf{f}, \mathbf{L} = \mathbf{l})$ have minima and maxima occurring at the same $\mathbf{f}$ and $\mathbf{l}$ configurations. In Equation 5.9, we have obtained a potential function, ϕ_{MP}, that is a smooth approximation of the broken parabola ϕ^*, by integrating out the line processes from the prior $\Pr(\mathbf{F} = \mathbf{f}, \mathbf{L} = \mathbf{l})$. Here, we evaluate the marginal intensity prior distribution of the temperature modified prior by integrating out the line processes and then show that the potential function arising from this marginal intensity is indeed $\phi_\beta(\Delta f_{ij})$. The marginal intensity distribution of the temperature modified prior is

$$
\begin{align}
\Pr_\beta(\mathbf{F}=\mathbf{f}) &= \sum_{\{\mathbf{l}\}} \Pr_\beta(\mathbf{F}=\mathbf{f},\mathbf{L}=\mathbf{l}) \tag{5.19}\\
&= \sum_{\{l^v_{ij}\},\{l^h_{ij}\}} \frac{1}{Z_6} e^{-\beta\lambda_2 \sum_{ij}\left\{[(\Delta^v f_{ij})^2(1-l^h_{ij})+\alpha l^h_{ij}]+[(\Delta^h f_{ij})^2(1-l^v_{ij})+\alpha l^v_{ij}]\right\}} \tag{5.20}\\
&= \frac{1}{Z_6}\prod_{ij}\left[\sum_{l^h_{ij}=\{0,1\}} e^{-\beta\lambda_2\{(\Delta^v f_{ij})^2(1-l^h_{ij})+\alpha l^h_{ij}\}}\right]\left[\sum_{l^v_{ij}=\{0,1\}} e^{-\beta\lambda_2\{(\Delta^h f_{ij})^2(1-l^v_{ij})+\alpha l^v_{ij}\}}\right] \notag\\
&= \frac{1}{Z_6}\prod_{ij}\left[e^{-\beta\lambda_2(\Delta^v f_{ij})^2}+e^{-\beta\lambda_2\alpha}\right]\left[e^{-\beta\lambda_2(\Delta^h f_{ij})^2}+e^{-\beta\lambda_2\alpha}\right] \tag{5.21}\\
&= \frac{1}{Z_6} e^{\sum_{ij}\left\{\log[e^{-\beta\lambda_2(\Delta^v f_{ij})^2}+e^{-\beta\lambda_2\alpha}]+\log[e^{-\beta\lambda_2(\Delta^h f_{ij})^2}+e^{-\beta\lambda_2\alpha}]\right\}} \notag\\
&= \frac{1}{Z_6} e^{-\beta\left\{\sum_{ij}\frac{-1}{\beta}\log[e^{-\beta\lambda_2(\Delta^v f_{ij})^2}+e^{-\beta\lambda_2\alpha}]+\sum_{ij}\frac{-1}{\beta}\log[e^{-\beta\lambda_2(\Delta^h f_{ij})^2}+e^{-\beta\lambda_2\alpha}]\right\}} \notag\\
&\stackrel{\text{def}}{=} \frac{1}{Z_6} e^{-\beta\left\{\sum_{ij}\phi_\beta(\Delta^v f_{ij})+\sum_{ij}\phi_\beta(\Delta^h f_{ij})\right\}} \tag{5.22}\\
&\stackrel{\text{def}}{=} \frac{1}{Z_6} e^{-\beta E_{MP}(\mathbf{f};\beta)} \notag
\end{align}
$$

where

$$E_{MP}(\mathbf{f};\beta) \stackrel{\text{def}}{=} \sum_{ij}\phi_\beta(\Delta^v f_{ij}) + \sum_{ij}\phi_\beta(\Delta^h f_{ij}).$$

The above steps are similar to Equations 5.10–5.14. We have shown that we can indeed obtain the potential function $\phi_\beta(\Delta f_{ij})$ in Equation 5.16 from the temperature exponentiated marginal intensity prior distribution, $\Pr_\beta(\mathbf{F}=\mathbf{f})$, where $\beta > 0$. Our deterministic annealing parameter, β, ranges from low to high values according to the annealing schedule.

In summary, the broken parabola, ϕ^*, was obtained by eliminating the line process from the original prior energy function in Section 5.2. This was done by minimizing the objective function with respect to the line process and then substituting the solution back into the objective function. In Section 5.2.1, the marginal prior distribution of the intensities was obtained by summing over all configurations of the line processes in the prior distribution. Following this, we showed the resulting marginal intensity prior was a smooth Gibbs prior and its potential function was a smooth approximation of the broken parabola. This smooth potential motivated us, in Section 5.2.2, to use a continuation method where a sequence of smooth potential functions, ϕ_β, indexed by a temperature parameter, β, was generated. The function, ϕ_β, was a smooth approximation of the broken parabola at high temperatures (low values of β) and approached the broken parabola at low temperatures. We showed that ϕ_β could also be obtained from a temperature modified marginal prior distribution which was obtained by introducing β to the original prior distribution $\Pr(\mathbf{F} = \mathbf{f}, \mathbf{L} = \mathbf{l})$. Integrating out the line processes, we obtained a marginal intensity prior distribution whose Gibbs potential function was indeed ϕ_β.

Our overall objective function to be minimized at each β is

$$\begin{aligned} E_\beta(\mathbf{f}) &= D(\mathbf{f}) + E_{MP}(\mathbf{f};\beta) \qquad (5.23) \\ &= D(\mathbf{f}) + \sum_{ij} \left[\phi_\beta(\Delta^v f_{ij}) + \phi_\beta(\Delta^h f_{ij})\right] \end{aligned}$$

$$
\begin{aligned}
&= \sum_{t\theta}\sum_{ij}\mathcal{H}_{t\theta;ij}f_{ij} - \sum_{t\theta} g_{t\theta}\log(\sum_{ij}\mathcal{H}_{t\theta;ij}f_{ij}) \\
&\quad + \left(\frac{-1}{\beta}\right)\sum_{ij}\left[\log\left(e^{-\beta\lambda_2(\Delta^v f_{ij})^2} + e^{-\beta\lambda_2\alpha}\right) + \log\left(e^{-\beta\lambda_2(\Delta^v f_{ij})^2} + e^{-\beta\lambda_2\alpha}\right)\right].
\end{aligned}
$$

Note that the likelihood objective function, $D(\mathbf{f})$, in the above is the same as in the original objective function in Equation 5.6 and does not depend on β. Since $\phi_\beta(\Delta f_{ij})$ approaches $\phi^*(\Delta f_{ij})$ at high values of β, the minimum of the above objective function approaches a minimum of the objective function in Equation 5.6 at high values of β.

Despite the absence of the line processes from our final objective function in Equation 5.23, line processes are implicitly contained in the broken parabola which is reached at limiting values of the control parameter β (Equation 5.5). This is because the marginal intensity prior originated from the prior distribution on $\mathbf{f}$ and $\mathbf{l}$. The property of line processes does not vanish when the line processes are integrated out, but is instead contained in the marginal distribution of the intensities. In the next subsection, we show that the property of line process is embodied in our objective function of Equation 5.23.

5.2.3 Conditional Mean Values of Line Processes in the Marginal Intensity Prior

Here, we show that the expected values of line processes at a given temperature can be obtained from the temperature exponentiated marginal intensity prior. The following is the derivation of the mean values of line processes conditioned on $\mathbf{f}$ and g at each β:

$$\begin{aligned} & E\{L^h_{ij} | \mathbf{F} = \mathbf{f}, \mathbf{G} = \mathbf{g}; \beta\} \\ = \; & 1 \cdot \mathrm{Pr}_\beta(L^h_{ij} = 1 | \mathbf{F} = \mathbf{f}, \mathbf{G} = \mathbf{g}) + 0 \cdot \mathrm{Pr}_\beta(L^h_{ij} = 0 | \mathbf{F} = \mathbf{f}, \mathbf{G} = \mathbf{g}) \\ = \; & 1 \cdot \mathrm{Pr}_\beta(L^h_{ij} = 1 | \mathbf{F} = \mathbf{f}) + 0 \cdot \mathrm{Pr}_\beta(L^h_{ij} = 0 | \mathbf{F} = \mathbf{f}) \\ = \; & \mathrm{Pr}_\beta(L_{ij} = 1 | \mathbf{F} = \mathbf{f}) \end{aligned}$$

Since the line processes can be completely described by $\mathbf{f}$ alone and independently of $\mathbf{g}$, the conditional probability $\mathrm{Pr}_\beta(L^h_{ij} = 1 | \mathbf{F} = \mathbf{f}, \mathbf{G} = \mathbf{g})$ becomes $\mathrm{Pr}_\beta(L^h_{ij} = 1 | \mathbf{F} = \mathbf{f})$. The conditional probability of the line processes at each β can be expressed by using Equations 5.18 and 5.21 in the following:

$$\begin{aligned} \mathrm{Pr}_\beta(\mathbf{L} = \mathbf{l} | \mathbf{F} = \mathbf{f}) &= \frac{\mathrm{Pr}_\beta(\mathbf{F} = \mathbf{f}, \mathbf{L} = \mathbf{l})}{\mathrm{Pr}_\beta(\mathbf{F} = \mathbf{f})} \\ &= \frac{(\frac{1}{Z_6}) e^{-\beta\lambda_2 \sum_{ij} \left[(\Delta^v f_{ij})^2 (1 - l^h_{ij}) + \alpha l^h_{ij} + (\Delta^h f_{ij})^2 (1 - l^v_{ij}) + \alpha l^v_{ij}\right]}}{(\frac{1}{Z_6}) \prod_{ij} \left[e^{-\beta\lambda_2\alpha} + e^{-\beta\lambda_2 (\Delta^v f_{ij})^2}\right] \left[e^{-\beta\lambda_2\alpha} + e^{-\beta\lambda_2 (\Delta^h f_{ij})^2}\right]} \\ &= \prod_{ij} \frac{e^{-\beta\lambda_2 \left[(\Delta^v f_{ij})^2 (1 - l^h_{ij}) + \alpha l^h_{ij}\right]} e^{-\beta\lambda_2 \left[(\Delta^h f_{ij})^2 (1 - l^v_{ij}) + \alpha l^v_{ij}\right]}}{\left[e^{-\beta\lambda_2\alpha} + e^{-\beta\lambda_2 (\Delta^v f_{ij})^2}\right] \left[e^{-\beta\lambda_2\alpha} + e^{-\beta\lambda_2 (\Delta^h f_{ij})^2}\right]} \end{aligned}$$

Because the line processes are independent of each other in our model, their mean values can be easily seen from the above conditional probability to be

$$\begin{aligned} E\{L^h_{ij} | \mathbf{F} = \mathbf{f}, \mathbf{G} = \mathbf{g}; \beta\} &= \mathrm{Pr}_\beta(L^h_{ij} = 1 | \mathbf{F} = \mathbf{f}) \\ &= \frac{e^{-\beta\lambda_2\alpha}}{e^{-\beta\lambda_2\alpha} + e^{-\beta\lambda_2 (\Delta^v f_{ij})^2}} \end{aligned}$$

$$= \frac{1}{1+e^{-\beta\lambda_2[(\Delta^v f_{ij})^2-\alpha]}} \tag{5.24}$$

and similarly,

$$E\{L_{ij}^v|\mathbf{F}=\mathbf{f},\mathbf{G}=\mathbf{g};\beta\} = \frac{1}{1+e^{-\beta\lambda_2[(\Delta^h f_{ij})^2-\alpha]}}. \tag{5.25}$$

Even though our objective function $E_\beta(\mathbf{f})$ in Equation 5.17 depends only on the intensities, it implicitly contains the conditional mean values of the line processes: we can always calculate the conditional mean values of the line processes from the intensity values alone. Since the temperature exponentiated prior eventually becomes the original prior distribution as the temperature is lowered from high values to zero, the conditional mean values of the line processes obtained from the temperature modified distribution eventually approach the "true" values of the line processes.

5.3 Derivation of the Deterministic Annealing Algorithm

Our original MAP estimate problem is now reduced to a combination of minimizing the objective function, $E_\beta(\mathbf{f})$, in Equation 5.17 at each β, and tracking the minimum through increasing values of β. A standard gradient descent minimization technique can be inefficient for an objective function such as $E_\beta(\mathbf{f})$ because of its global connectivity. Instead, we use the incomplete/complete data formulation for the EM approach discussed in Section 3.2. Following Equations 3.23 and 3.24, the EM algorithm in our case can be written

$$\begin{aligned}
\text{E-step : Form} \quad & Q(\mathbf{f}|\hat{\mathbf{f}}^n) = E_{\mathbf{C}}\{\log \Pr(\mathbf{C}|\mathbf{F}=\mathbf{f})|\mathbf{G}=\mathbf{g}, \mathbf{F}=\hat{\mathbf{f}}^n\} \\
\text{M-step : Solve} \quad & \arg\max_{\mathbf{f}} \left[Q(\mathbf{f}|\hat{\mathbf{f}}^n) - E_{MP}(\mathbf{f};\beta)\right] \\
& = \arg\min_{\mathbf{f}} \left[-Q(\mathbf{f}|\hat{\mathbf{f}}^n) + E_{MP}(\mathbf{f};\beta)\right]. \qquad (5.26)
\end{aligned}$$

Recall Equation 3.16 for $Q(\mathbf{f}|\hat{\mathbf{f}}^n)$. Note that we formulated our optimization problem as an equivalent minimization problem instead of a maximization problem. The M-step objective function to be minimized is

$$\begin{aligned}
M(\mathbf{f}|\hat{\mathbf{f}}^n;\beta) &\overset{\text{def}}{=} -Q(\mathbf{f}|\hat{\mathbf{f}}^n) + E_{MP}(\mathbf{f};\beta) \\
&= -Q(\mathbf{f}|\hat{\mathbf{f}}^n) + \sum_{ij}\left[\phi_\beta(\Delta^h f_{ij}) + \phi_\beta(\Delta^v f_{ij})\right] \\
&= \sum_{ij}\sum_{t\theta} \mathcal{H}_{t\theta,ij} f_{ij} - \sum_{ij}\sum_{t\theta} g_{t\theta} \frac{\mathcal{H}_{t\theta;ij}\hat{f}^n_{ij}}{\sum_{kl}\mathcal{H}_{t\theta;kl}\hat{f}^n_{kl}} \log(f_{ij}) \\
&\quad + \sum_{ij}\left(-\frac{1}{\beta}\right)\log\left[e^{-\beta\lambda_2\alpha} + e^{-\beta\lambda_2(\Delta^v f_{ij})^2}\right] \\
&\quad + \sum_{ij}\left(-\frac{1}{\beta}\right)\log\left[e^{-\beta\lambda_2\alpha} + e^{-\beta\lambda_2(\Delta^h f_{ij})^2}\right]. \qquad (5.27)
\end{aligned}$$

Thus, $E_\beta(\mathbf{f}) = D(\mathbf{f}) + E_{MP}(\mathbf{f};\beta)$ has been transformed, via the EM approach, to the minimization of $M(\mathbf{f}|\hat{\mathbf{f}}^n;\beta)$ at each EM iteration at a given β. The transformation of $D(\mathbf{f})$ to $-Q(\mathbf{f}|\hat{\mathbf{f}}_n)$ is shown in Appendix B. Due to the prior, however, the M-step objective is difficult to minimize (see Section 3.2.2). Instead we use a coordinate-wise descent technique for the minimization. This, however, only decreases the M-step objective function at each EM iteration rather than minimizing it, which is the same as the generalized EM (GEM) algorithm discussed in Section 3.2.2.1. When we are at location (i,j), the objective function is minimized with respect to $\mathbf{f}$ at (i,j) keeping all other variables fixed. A new location is then chosen, and the method is repeated. After a full sweep of the lattice, the procedure is repeated until convergence criteria (typically objective differences or norms) are met. This method *always* converges to a local minimum [36]. This local minimum is then used as the initial estimate for the next temperature setting and a local minimum at this new temperature is then found. By varying the temperature in this way, we attempt to find the global minimum. Unlike gradient descent methods, the coordinate-wise descent technique also avoids concerns of choice of step sizes. The closed form solution for each variable can be obtained simply by differentiating the objective function and solving for the variable of interest. Hence,

$$\frac{\partial M(\mathbf{f}|\hat{\mathbf{f}}^n)}{\partial f_{ij}} = -\frac{\partial Q(\mathbf{f}|\hat{\mathbf{f}}^n)}{\partial f_{ij}} + \frac{\partial E_{MP}(\mathbf{f})}{\partial f_{ij}} \tag{5.28}$$

where $\frac{\partial Q(\mathbf{f}|\hat{\mathbf{f}}^n)}{\partial f_{ij}}$ was shown in Equation 3.17 to be

$$-\frac{\partial Q(\mathbf{f}|\hat{\mathbf{f}}^n)}{\partial f_{ij}} = \sum_{t\theta} \mathcal{H}_{t\theta;ij} - \sum_{t\theta} g_{t\theta} \frac{\mathcal{H}_{t\theta;ij}\hat{f}^n_{ij}}{\sum_{kl} \mathcal{H}_{t\theta;kl}\hat{f}^n_{kl}} \frac{1}{f_{ij}},$$

and,

$$\begin{aligned}\frac{\partial E_{MP}(\mathbf{f})}{\partial f_{ij}} = -\frac{1}{\beta}\Bigg[& \frac{e^{-\beta\lambda_2(\Delta^v f_{ij})^2}}{e^{-\beta\lambda_2\alpha}+e^{-\beta\lambda_2(\Delta^v f_{ij})^2}}\left[-2\beta\lambda_2(f_{i+1,j}-f_{ij})(-1)\right] \\ & +\frac{e^{-\beta\lambda_2(\Delta^v f_{i-1,j})^2}}{e^{-\beta\lambda_2\alpha}+e^{-\beta\lambda_2(\Delta^v f_{i-1,j})^2}}\left[-2\beta\lambda_2(f_{ij}-f_{i-1,j})\right] \\ & +\frac{e^{-\beta\lambda_2(\Delta^h f_{ij})^2}}{e^{-\beta\lambda_2\alpha}+e^{-\beta\lambda_2(\Delta^h f_{ij})^2}}\left[-2\beta\lambda_2(f_{i,j+1}-f_{ij})(-1)\right] \\ & +\frac{e^{-\beta\lambda_2(\Delta^h f_{i,j-1})^2}}{e^{-\beta\lambda_2\alpha}+e^{-\beta\lambda_2(\Delta^h f_{i,j-1})^2}}\left[-2\beta\lambda_2(f_{ij}-f_{i,j-1})\right]\Bigg]. \end{aligned} \tag{5.29}$$

We are interested in obtaining a closed-form solution for f_{ij} from $\frac{\partial M(\mathbf{f}|\hat{\mathbf{f}}^n;\beta)}{\partial f_{ij}} = 0$ at each location, while keeping all other $\mathbf{f}$ variables fixed. However, upon examining Equation 5.29, it is clear that such an explicit solution is impossible due to the transcendental nature of the exponential term. One possible solution to this problem is to separate the global dependence of the variables by introducing a new, albeit dependent (on $\mathbf{f}$) variable z_{ij} [4] and to descend on each z_{ij} and f_{ij} separately. Let us first consider one of the exponential terms in Equation 5.29, $\frac{e^{-\beta\lambda_2(\Delta^v f_{ij})^2}}{e^{-\beta\lambda_2\alpha}+e^{-\beta\lambda_2(\Delta^v f_{ij})^2}}$. By examining the mean value of the horizontal line process at (i,j) in Equation 5.24, the exponential term is equivalent to

$$1 - \text{conditional mean value of } l^h_{ij}.$$

This suggests we take z_{ij} to be the conditional mean value of line process, l_{ij}.

We now define

$$\begin{aligned} z^v_{ij} &\overset{\text{def}}{=} \frac{e^{-\beta\lambda_2\alpha}}{e^{-\beta\lambda_2\alpha}+e^{-\beta\lambda_2(\Delta^h f_{ij})^2}} = \frac{1}{1+e^{-\beta\lambda_2[(\Delta^h f_{ij})^2-\alpha]}} \\ z^h_{ij} &\overset{\text{def}}{=} \frac{e^{-\beta\lambda_2\alpha}}{e^{-\beta\lambda_2\alpha}+e^{-\beta\lambda_2(\Delta^v f_{ij})^2}} = \frac{1}{1+e^{-\beta\lambda_2[(\Delta^v f_{ij})^2-\alpha]}}. \end{aligned} \tag{5.30}$$

From the definition of z, we then have

$$\begin{aligned} 1-z^v_{ij} &= \frac{e^{-\beta\lambda_2(\Delta^h f_{ij})^2}}{e^{-\beta\lambda_2\alpha}+e^{-\beta\lambda_2(\Delta^h f_{ij})^2}} = \frac{1}{1+e^{-\beta\lambda_2[\alpha-(\Delta^h f_{ij})^2]}} \\ 1-z^h_{ij} &= \frac{e^{-\beta\lambda_2(\Delta^v f_{ij})^2}}{e^{-\beta\lambda_2\alpha}+e^{-\beta\lambda_2(\Delta^v f_{ij})^2}} = \frac{1}{1+e^{-\beta\lambda_2[\alpha-(\Delta^v f_{ij})^2]}}. \end{aligned} \tag{5.31}$$

However, arbitrarily changing variables in the objective function and then performing 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. In the following section (Section 5.4), we show that such a corresponding objective function of **f** and **z** can be derived by using Lagrange parameters. This corresponding function is

$$\begin{aligned} M^\circ(\mathbf{f},\mathbf{z}|\hat{\mathbf{f}}^n;\beta) = {} & \sum_{ij}\sum_{t\theta}\mathcal{H}_{t\theta;ij}f_{ij} - \sum_{ij}\sum_{t\theta} g_{t\theta}\frac{\mathcal{H}_{t\theta;ij}\hat{f}^n_{ij}}{\sum_{kl}\mathcal{H}_{t\theta;kl}\hat{f}^n_{kl}}\log(f_{ij}) \\ & +\sum_{ij}\lambda_2\left[(\Delta^v f_{ij})^2(1-z^h_{ij})+\alpha z^h_{ij}\right] + \sum_{ij}\frac{1}{\beta}\left[z^h_{ij}\log(z^h_{ij})+(1-z^h_{ij})\log(1-z^h_{ij})\right] \\ & +\sum_{ij}\lambda_2\left[(\Delta^h f_{ij})^2(1-z^v_{ij})+\alpha z^v_{ij}\right] + \sum_{ij}\frac{1}{\beta}\left[z^v_{ij}\log(z^v_{ij})+(1-z^v_{ij})\log(1-z^v_{ij})\right]. \end{aligned} \tag{5.32}$$

Thus minimizing the above function, $M^\circ(\mathbf{f},\mathbf{z}|\hat{\mathbf{f}}^n;\beta)$, gives us the same answer as minimizing $M(\mathbf{f}|\hat{\mathbf{f}}^n;\beta)$. We perform coordinate-wise descent on **f** while keeping **z** constant, and on $\mathbf{z}$ while $\mathbf{f}$ is held constant. Now the M-step in our GEM algorithm becomes two steps:

$$\begin{array}{lll} \text{M1-step :} & \text{Fix } \mathbf{z} \text{ values, } \dfrac{\partial M^{\circ}(\mathbf{f},\mathbf{z}|\hat{\mathbf{f}}^{n};\beta)}{\partial f_{ij}} = 0 & \Rightarrow \quad \hat{f}_{ij}^{n+1} \text{ for all } i,j \\ \text{M2-step :} & \text{Fix } \mathbf{f} \text{ values, } \dfrac{\partial M^{\circ}(\mathbf{f},\mathbf{z}|\hat{\mathbf{f}}^{n};\beta)}{\partial z_{ij}} = 0 & \Rightarrow \quad \hat{z}_{ij}^{n+1} \text{ for all } i,j \end{array}$$

M1-step : We take the partial derivative of the objective function $M^{\circ}(\mathbf{f},\mathbf{z}|\hat{\mathbf{f}}^{n};\beta)$ in Equation 5.32 with respect to f_{ij} and get

$$\begin{aligned} \frac{\partial M^{\circ}(\mathbf{f},\mathbf{z}|\hat{\mathbf{f}}^{n};\beta)}{\partial f_{ij}} &= \sum_{t\theta} \mathcal{H}_{t\theta;ij} - \sum_{t\theta} g_{t\theta} \frac{\mathcal{H}_{t\theta;ij} \hat{f}_{ij}^{n}}{\sum_{kl} \mathcal{H}_{t\theta;kl} \hat{f}_{kl}^{n}} \frac{1}{f_{ij}} \\ &+2\lambda_2 f_{ij} \left[(1-z_{ij}^{h}) + (1-z_{i-1,j}^{h}) + (1-z_{ij}^{v}) + (1-z_{i,j-1}^{v})\right] \\ &-2\lambda_2 \left[f_{i+1,j}(1-z_{ij}^{h}) + f_{i-1,j}(1-z_{i-1,j}^{h}) + f_{i,j+1}(1-z_{ij}^{v}) + f_{i,j-1}(1-z_{i,j-1}^{v})\right]. \end{aligned} \tag{5.33}$$

Note that if we use the definition of $\mathbf{z}$ in Equation 5.30 and substitute it in Equation 5.29, the original M-step of Equation 5.28 becomes the same as Equation 5.33. This shows the consistency of our transformed M-step function (Equation 5.32). If we set Equation 5.33 to zero, we get a quadratic equation in f_{ij}. Solving this equation for f_{ij} and choosing the non-negative root, we get an update equation for f_{ij} at GEM iteration $n+1$:

$$\hat{f}_{ij}^{n+1} = \frac{-(\sum_{t\theta} \mathcal{H}_{t\theta;ij} - 2\lambda_2 X_3) + \sqrt{(\sum_{t\theta} \mathcal{H}_{t\theta;ij} - 2\lambda_2 X_3)^2 + 8\lambda_2 X_2 X_1}}{4\lambda_2 X_2}, \tag{5.34}$$

where

$$\begin{aligned}
X_1 &\stackrel{\mathrm{def}}{=} \sum_{t\theta} g_{t\theta} \frac{\mathcal{H}_{t\theta;ij}\hat{f}^n_{ij}}{\sum_{kl}\mathcal{H}_{t\theta;kl}\hat{f}^n_{kl}} \\
X_2 &\stackrel{\mathrm{def}}{=} (1 - z^h_{ij}\,|_{\hat{\mathbf{f}}^n}) + (1 - z^h_{i-1,j}\,|_{\hat{\mathbf{f}}^n}) + (1 - z^v_{ij}\,|_{\hat{\mathbf{f}}^n}) + (1 - z^v_{i,j-1}\,|_{\hat{\mathbf{f}}^n}) \\
X_3 &\stackrel{\mathrm{def}}{=} f_{i+1,j}(1 - z^h_{ij}\,|_{\hat{\mathbf{f}}^n}) + f_{i-1,j}(1 - z^h_{i-1,j}\,|_{\hat{\mathbf{f}}^n}) \\
&+ f_{i,j+1}(1 - z^v_{ij}\,|_{\hat{\mathbf{f}}^n}) + f_{i,j-1}(1 - z^v_{i,j-1}\,|_{\hat{\mathbf{f}}^n}).
\end{aligned}$$

The notation $z^v_{ij}\,|_{\hat{\mathbf{f}}^n}$ stands for z^v_{ij} evaluated using $\hat{\mathbf{f}}^n$, and similarly for $z^h_{ij}\,|_{\hat{\mathbf{f}}^n}$. The value of X_2 is non-zero except when all the surrounding line processes of a pixel, f_{ij}, are exactly equal to one. For the case when all the surrounding line processes are equal to one, Equation 5.33 reduces to $-\frac{\partial Q}{\partial f_{ij}}$ of the EM–ML method (see Equation 3.17). Our temperature varying scheme, however, avoids the creation of extreme hot spots. Since we update pixels sequentially by using the raster scan, we use the most recent estimates of $f_{i\pm1,j\pm1}$ for computing X_3. The intensity update equation also guarantees positivity of f_{ij} because of the non-negativity of X_1, X_2 and X_3.

M2-step : For the update equation for z_{ij}, we take partial derivatives of Equation 5.32 with respect to z^v_{ij} and z^h_{ij}, and set the equations to zero to solve for z^h_{ij} and z^v_{ij}. The following shows the steps used in solving for z^v_{ij}:

$$\frac{\partial M^\circ(\mathbf{f},\mathbf{z}|\hat{\mathbf{f}}^n;\beta)}{\partial z^v_{ij}} = -\lambda_2(\Delta^h f_{ij})^2 + \lambda_2\alpha + \frac{1}{\beta}[\log(z^v_{ij}) - \log(1 - z^v_{ij})] = 0$$

$$
\begin{aligned}
\Rightarrow \quad & \log(\frac{1-z^v_{ij}}{z^v_{ij}}) = -\beta\lambda_2(\Delta^h f_{ij})^2 + \beta\lambda_2\alpha \\
\Rightarrow \quad & \frac{1}{z^v_{ij}} - 1 = e^{-\beta[\lambda_2(\Delta^h f_{ij})^2 - \lambda_2\alpha]} \\
\Rightarrow \quad & z^v_{ij} = \frac{1}{1 + e^{-\beta[\lambda_2(\Delta^h f_{ij})^2 - \lambda_2\alpha]}}.
\end{aligned}
$$

The update equation obtained from the coordinate descent method for z^v_{ij} is the same as our definition in Equation 5.30, as expected. This verifies the consistency of our definition. The update equation for $\mathbf{z}$ at GEM iteration $n+1$ can, hence, be written

$$
\begin{aligned}
z^v_{ij}\ \big|_{\hat{\mathbf{f}}^{n+1}} &= \frac{1}{1+e^{-\beta[\lambda_2(\Delta^h f^{n+1}_{ij})^2 - \lambda_2\alpha]}}, \\
z^h_{ij}\ \big|_{\hat{\mathbf{f}}^{n+1}} &= \frac{1}{1+e^{-\beta[\lambda_2(\Delta^v f^{n+1}_{ij})^2 - \lambda_2\alpha]}}.
\end{aligned}
\tag{5.35}
$$

To ensure that the solution obtained by differentiating M° with respect to f_{ij} and to z_{ij} and setting it to zero is indeed a minimum, it is sufficient to show that M° is separately convex with respect to f_{ij} and z_{ij}. This can be done by examining the second derivatives:

$$
\begin{aligned}
& \frac{\partial^2 M^\circ(\mathbf{f}, \mathbf{z}|\hat{\mathbf{f}}^n; \beta)}{\partial f^2_{ij}} \\
& \quad = \frac{1}{f^2_{ij}} \left(\sum_{t\theta} g_{t\theta} \frac{\mathcal{H}_{t\theta;ij} \hat{f}^n_{ij}}{\sum_{kl} \mathcal{H}_{t\theta;kl} \hat{f}^n_{kl}} \right) + 2\lambda_2 \left(4 - z^h_{ij} - z^h_{i-1,j} - z^v_{ij} - z^v_{i,j-1}\right) \quad \geq \quad 0
\end{aligned}
$$

$$
\frac{\partial^2 M^\circ(\mathbf{f}, \mathbf{z}|\hat{\mathbf{f}}^n; \beta)}{\partial (z^v_{ij})^2} = \frac{1}{\beta z^v_{ij}(1 - z^v_{ij})} \quad \geq \quad 0.
$$

The second derivative, $\frac{\partial^2 M^\circ}{\partial f_{ij}^2}$, goes to zero only when $\hat{f}_{ij}^n$ is zero and all four surrounding z's are exactly equal to one. The second derivative, $\frac{\partial^2 M^\circ}{\partial (z_{ij}^v)^2}$, goes to zero when $\beta \rightarrow \infty$. In practice, however, β never reaches infinity, therefore z never becomes exactly zero or one. We have guaranteed that a single M-step taken in any direction is a descent step since the fixed point in that direction corresponds to a minimum. This minimum is achieved in a single step. Hence, convergence of the GEM procedure to a local minimum of M° is guaranteed.

Our deterministic annealing algorithm can be outlined as follows:

```
do {
    do {
        /* GEM procedure */
        $\hat{\mathbf{f}}^n \longrightarrow \hat{\mathbf{f}}^{n+1}$
        $\hat{\mathbf{z}}^n \longrightarrow \hat{\mathbf{z}}^{n+1}$
    }while$\Big( \Big| M(\hat{\mathbf{f}}^{n+1}, \hat{\mathbf{z}}^{n+1} | \hat{\mathbf{f}}^n; \beta) - M(\hat{\mathbf{f}}^n, \hat{\mathbf{z}}^n | \hat{\mathbf{f}}^n; \beta) \Big| > \epsilon \Big)$
    increase $\beta$
}while( z values not converged to either 0 or 1 within a tolerance)
```

We start with a high value of temperature (low β), and at each temperature carry out the GEM procedure until convergence. Convergence is reached when the difference of consecutive M-step objectives is within some convergence parameter, ϵ. We then lower the temperature (increase β) and carry out the GEM procedure

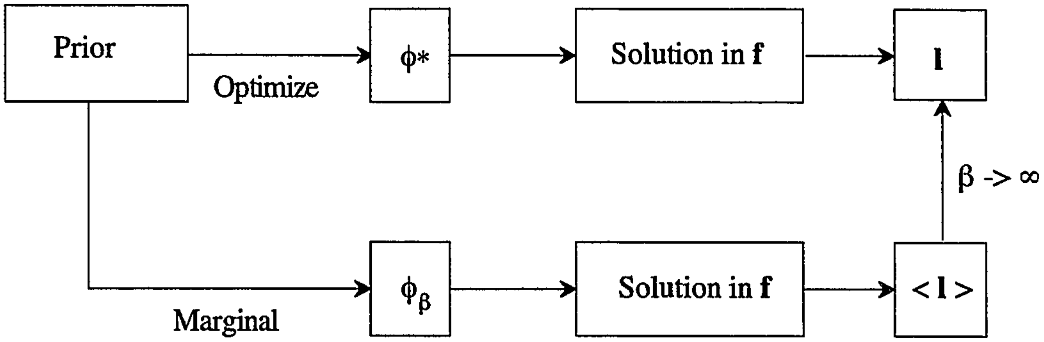


Figure 5.4:

at the new temperature. The temperature is lowered until all mean values of line processes, z, converge to either 0 or 1 within a tolerance. At each temperature, the GEM procedure consists of two kinds of updates. We first update **f**, according to the update equation Equation 5.34, over the entire image with z values held fixed, and then update z, according to Equation 5.35, over the entire image using the new **f** values. Once the new estimate, $\hat{\mathbf{f}}^{n+1}$, has been obtained, a new M-step objective function is obtained by substituting $\hat{\mathbf{f}}^{n+1}$ for $\hat{\mathbf{f}}^{n}$ to get $M(\mathbf{f}, \mathbf{z}|\hat{\mathbf{f}}^{n+1}; \beta)$. In this manner, the entire GEM procedure is carried out until convergence. Figure 5.4 shows the relation of the temperature modified marginal prior to the original prior, ϕ^*, in the deterministic annealing method.

5.4 Convergence to a Local Minimum in the Deterministic Annealing Algorithm

As mentioned in the previous section (Section 5.3), we do not yet know whether coordinate-wise descent on the new variable $\mathbf{z}$ in our objective function is guaranteed to converge. If it does converge, is it in fact a local minimum rather than a maximum or a saddle point? The convergence can be proved by producing an objective function that corresponds to the new variables, $\mathbf{z}$. By examining the first and second order derivatives, we also show that a local minimum is indeed achieved.

In this section, we start with the M-step function in Equation 5.27 and systematically derive the objective function (Equation 5.32) corresponding to the change of variables $\mathbf{z} = \mathbf{z}(\mathbf{f})$. Recall the M-step objective function in Equation 5.27:

$$\begin{aligned}
M(\mathbf{f}|\hat{\mathbf{f}}^n;\beta) &\stackrel{\text{def}}{=} -Q(\mathbf{f}|\hat{\mathbf{f}}^n) + E_{\text{MP}}(\mathbf{f};\beta) \\
&= -Q(\mathbf{f}|\hat{\mathbf{f}}^n) + \sum_{ij}\left[\phi_\beta(\Delta^h f_{ij}) + \phi_\beta(\Delta^v f_{ij})\right] \\
&= \sum_{ij}\sum_{t\theta}\mathcal{H}_{t\theta,ij}f_{ij} - \sum_{ij}\sum_{t\theta} g_{t\theta}\frac{\mathcal{H}_{t\theta;ij}\hat{f}^n_{ij}}{\sum_{kl}\mathcal{H}_{t\theta;kl}\hat{f}^n_{kl}}\log(f_{ij}) \\
&\quad + \sum_{ij}\left(-\frac{1}{\beta}\right)\log\left[e^{-\beta\lambda_2\alpha} + e^{-\beta\lambda_2(\Delta^v f_{ij})^2}\right] \\
&\quad + \sum_{ij}\left(-\frac{1}{\beta}\right)\log\left[e^{-\beta\lambda_2\alpha} + e^{-\beta\lambda_2(\Delta^h f_{ij})^2}\right]
\end{aligned}$$

The work in this section is done in close conjunction with the work of Anand Rangarajan

A change of variables in the objective function can be performed by the method of Lagrange parameters. Consider the following transformation to be applied at every location (i, j),

$$\psi(\Delta^h f_{ij}, s^v_{ij}) \stackrel{\text{def}}{=} \xi(s^v_{ij}) + \mu^v_{ij}((\Delta^h f_{ij})^2 - s^v_{ij}), \tag{5.36}$$

where a similar relationship applies for $\Delta^v f_{ij}$, s^h_{ij} and μ^h_{ij}. In Equation 5.36,

$$\xi(s^v_{ij}) \stackrel{\text{def}}{=} -\frac{1}{\beta} \log \left[e^{-\beta\lambda_2\alpha} + e^{-\beta\lambda_2 s^v_{ij}} \right],$$

with μ^v_{ij} being a Lagrange parameter (corresponding to s^v_{ij}) which will be determined shortly. Note that the definition of $\xi(s^v_{ij})$ is written in terms of a dummy variable, s^v_{ij}. The variable s^v is forced towards the value $(\Delta^h f_{ij})^2$ by the action of the Lagrange constraint term in Equation 5.36 (then $\psi(\Delta^h f_{ij}, s^v_{ij})$ and $\psi(\Delta^v f_{ij}, s^h_{ij})$ become $\phi_\beta(\Delta^h f_{ij})$ and $\phi_\beta(\Delta^h f_{ij})$, respectively). Differentiating $\xi(s^v_{ij})$ with respect to s^v_{ij} and setting the result to zero, we get

$$\mu^v_{ij} = \xi'(s^v_{ij}) = \frac{\lambda_2 e^{-\beta\lambda_2 s^v_{ij}}}{e^{-\beta\lambda_2\alpha} + e^{-\beta\lambda_2 s^v_{ij}}} = \frac{\lambda_2 e^{-\beta\lambda_2(s^v_{ij}-\alpha)}}{1 + e^{-\beta\lambda_2(s^v_{ij}-\alpha)}}.$$

The transformation after Lagrange parameter elimination is

$$\psi(\Delta^h f_{ij}, s^v_{ij}) = \xi(s^v_{ij}) + \xi'(s^v_{ij})((\Delta^h f_{ij})^2 - s^v_{ij}). \tag{5.37}$$

In Equation 5.37, when the derivative with respect to s^v_{ij} is set to zero, we get

$$\begin{aligned}
\frac{\partial \psi(\Delta^h f_{ij}, s^v_{ij})}{\partial s^v_{ij}} = 0 \;&\Rightarrow\; \xi'(s^v_{ij}) + \xi''(s^v_{ij})((\Delta^h f_{ij})^2 - s^v_{ij}) - \xi'(s^v_{ij}) = 0 \\
&\Rightarrow\; \xi''(s^v_{ij})((\Delta^h f_{ij})^2 - s^v_{ij}) = 0. \qquad (5.38)
\end{aligned}$$

There are two solutions for the above equation. One is when $s^v_{ij} = (\Delta^h f_{ij})^2$ and the other is when $\xi''(s^v_{ij}) = 0$. The fixed point of interest is at $s^v_{ij} = (\Delta^h f_{ij})^2$. To ensure that this fixed point is a minimum, we examine the second partial derivative of $\psi(\Delta^h f_{ij}, s^v_{ij})$ with respect to s^v_{ij}:

$$\begin{aligned}
\frac{\partial^2 \psi(\Delta^h f_{ij}, s^v_{ij})}{\partial (s^v_{ij})^2} &= \frac{\partial \{\xi''(s^v_{ij})((\Delta^h f_{ij})^2 - s^v_{ij})\}}{\partial s^v_{ij}} \\
&= \xi'''(s^v_{ij})((\Delta^h f_{ij})^2 - s^v_{ij}) - \xi''(s^v_{ij}) \\
&\qquad\qquad \Downarrow \text{ evaluate at } s^v_{ij} = (\Delta^h f_{ij})^2 \\
&\quad -\xi''(s^v_{ij}),
\end{aligned}$$

where,

$$\xi''(s^v_{ij}) = -\beta\lambda_2^2 \frac{e^{-\beta\lambda_2(s^v_{ij}-\alpha)}}{\left[1 + e^{-\beta\lambda_2(s^v_{ij}-\alpha)}\right]^2}.$$

Notice that $\xi''(s^v_{ij}) < 0$, meeting the positivity condition of the second partial derivative of $\psi(\Delta^h f_{ij}, s^v_{ij})$ at the fixed point $s^v_{ij} = (\Delta^h f_{ij})^2$. The case $\xi''(s^v_{ij}) = 0$ occurs when either $\beta \rightarrow \infty$ or $s^v_{ij} \rightarrow \infty$. Since, in practice, β never reaches infinity and $s_{ij} \in [0, \infty)$, $\xi''(s^v_{ij}) = 0$ does not occur. Therefore, the fixed point $s^v_{ij} = (\Delta^h f_{ij})^2$ is a minimum. We now have an objective function in the two sets of variables f_{ij} and s_{ij} along with the fixed point condition $s^v_{ij} = (\Delta^h f_{ij})^2$ and $s^h_{ij} = (\Delta^v f_{ij})^2$.

We now rewrite Equation 5.37 as

$$\psi(\Delta^h f_{ij}, s^v_{ij}) = (\Delta^h f_{ij})^2 \xi'(s^v_{ij}) + \left(\xi(s^v_{ij}) - s^v_{ij}\xi'(s^v_{ij})\right).$$

Using the transformation $z^v_{ij} = z(s^v_{ij}) = 1 - \frac{\xi'(s^v_{ij})}{\lambda_2}$ and $z^h_{ij} = z(s^h_{ij}) = 1 - \frac{\xi'(s^h_{ij})}{\lambda_2}$ (note that this transformation is a restatement of (Equation 5.30)), we get

$$\begin{aligned}
\psi(\Delta^h f_{ij}, z^v_{ij}) &= \lambda_2(\Delta^h f_{ij})^2(1 - z^v_{ij}) + \lambda_2 \alpha z^v_{ij} \\
&\quad + \frac{1}{\beta}[z^v_{ij}\log(z^v_{ij}) + (1 - z^v_{ij})\log(1 - z^v_{ij})] \\
\psi(\Delta^v f_{ij}, z^h_{ij}) &= \lambda_2(\Delta^v f_{ij})^2(1 - z^h_{ij}) + \lambda_2 \alpha z^h_{ij} \\
&\quad + \frac{1}{\beta}[z^h_{ij}\log(z^h_{ij}) + (1 - z^h_{ij})\log(1 - z^h_{ij})].
\end{aligned} \tag{5.39}$$

Using Equation 5.39, we get the following corresponding objective function of $M(\mathbf{f}|\hat{\mathbf{f}}^n;\beta)$ in Equation 5.27

$$M^\circ(\mathbf{f},\mathbf{z}|\hat{\mathbf{f}}^n;\beta) = \sum_{ij}\sum_{t\theta} \mathcal{H}_{t\theta,ij} f_{ij} - \sum_{ij}\sum_{t\theta} g_{t\theta} \frac{\mathcal{H}_{t\theta;ij}\hat{f}^n_{ij}}{\sum_{kl}\mathcal{H}_{t\theta;kl}\hat{f}^n_{kl}} \log(f_{ij}) \tag{5.40}$$

$$+\sum_{ij} \lambda_2 \left[(\Delta^h f_{ij})^2(1 - z^v_{ij}) + \alpha z^v_{ij}\right] + \sum_{ij} \left(\frac{1}{\beta}\right) \left[z^v_{ij} \log(z^v_{ij}) + (1 - z^v_{ij}) \log(1 - z^v_{ij})\right]$$
$$+\sum_{ij} \lambda_2 \left[(\Delta^v f_{ij})^2(1 - z^h_{ij}) + \alpha z^h_{ij}\right] + \sum_{ij} \left(\frac{1}{\beta}\right) \left[z^h_{ij} \log(z^h_{ij}) + (1 - z^h_{ij}) \log(1 - z^h_{ij})\right].$$

Now we perform the coordinate descent on the above function M°. First we descend on z: we take a partial derivative of M° with respect to z^v_{ij} by fixing all other z^v_{kl} and all f_{kl} and then we set it equal to zero and solve for z^v_{ij} (similarly for z^v_{ij}).

$$\begin{aligned}
&\frac{\partial M^\circ(\mathbf{f}, \mathbf{z}|\hat{\mathbf{f}}^n; \beta)}{z^v_{ij}} = 0 \\
\Rightarrow \quad & -\lambda_2(\Delta^h f_{ij})^2 + \alpha + \frac{1}{\beta}\left[\log(z^v_{ij}) + 1\right] + \frac{1}{\beta}\left[-\log(1 - z^v_{ij}) - 1\right] = 0 \\
\Rightarrow \quad & \frac{z^v_{ij}}{1 - z^v_{ij}} = e^{\beta\lambda_2\left[(\Delta^h f_{ij})^2 - \alpha\right]} \\
\Rightarrow \quad & z^v_{ij} = \frac{1}{1 + e^{-\beta\lambda_2\left[(\Delta^h f_{ij})^2 - \alpha\right]}}.
\end{aligned}$$

Notice that the solution for z is the same as the conditional mean value of the line processes as we expect. If we substitute the solutions back into the function M°, we get our original M-step function $M(\mathbf{f}|\hat{\mathbf{f}}^n; \beta)$.

We have shown that there exists a corresponding objective function, M°, to the change of variable, $\mathbf{z}$, and that the coordinate-wise descent on the new variable, $\mathbf{z}$, indeed converges to a local minimum.

5.5 Related Work

Below, we discuss related work that utilize Bayesian models with priors to attempt to enforce piecewise smoothness. The priors are expressed through Gibbs potential functions. We can group these efforts into two types: those that use priors defined only with respect to the intensities, and those with priors using both intensities and line processes.

5.5.1 Prior Defined on Intensities

All of the following efforts used a Gibbs potential function, ϕ, defined only on intensities. There were no associated mixed variable objective functions from which the ϕ functions were derived. Most such reconstructions using Gibbs potential functions are improvements over the EM–ML reconstruction; however, some of the ϕ functions lack the property of allowing formation of discontinuities.

Geman and McClure [16] used a ϕ function of the form $\phi(\Delta f) = \frac{(\Delta f)^2}{1+(\Delta f)^2}$. They used stochastic relaxation for the MMSE estimation and a gradient descent method for the MAP estimation. In the context of image restoration, Geman and Reynolds [14] later mentioned the possible use of a more general ϕ function: $\phi(\Delta f) = \frac{|\Delta f|^k}{1+|\Delta f|^k}, k = 1, 2, \ldots$ Their results were obtained using stochastic relaxation with a ϕ function having $k = 1$. They use both the first order derivative ($\Delta f_i = f_i - f_{i-1}$), which is more effective in recovering the discontinuities, and the higher order derivatives (for example, the second order derivative, $\Delta^2 f_i = \Delta f_i - \Delta f_{i-1}$), which are supposedly more suitable for recovering the basic geometric structure of regions. This is implemented so that the first-order model generates a starting point for the second-order model, which in turn provides the starting point for the third-order model, and so on. In general, stochastic relaxation is very computationally demanding since the conditional distribution at each pixel site has to be computed for the full range of possible pixel values. Geman and McClure decreased the number of operations by selectively reducing the sampling range instead of using the full dynamic range.

Hebert and Leahy [20] used three different ϕ functions: (i) $\phi(\Delta f) = (\Delta f)^2$, which was earlier used by Geman and Geman [15] in image restoration, increasingly penalizes the difference between neighboring pixels; (ii) $\phi(\Delta f) = \frac{(\Delta f)^2}{1+(\Delta f)^2}$ suggested by Geman and McClure; and (iii) $\phi(\Delta f) = \log(1 + (\Delta f)^2)$ which is a compromise between (i) and (ii). Their main concern was to integrate a ϕ function with the likelihood into a GEM algorithm.

Green [19] suggested a convex ϕ function: $\phi(\Delta f) = \log(\cosh(\Delta f))$ and devised a one-step-late (OSL) algorithm for optimization. While having several mathematically desirable properties, Green's ϕ function does not promote the formation of discontinuities. The OSL algorithm is an EM-type approach in which the partial derivative of the prior energy function (equivalent to our $\frac{\partial E_{\mathrm{MP}}(\mathbf{f})}{\partial f_{ij}}$ in Equation 5.29) is approximated using the current estimate of $\mathbf{f}^n$ to get a closed form update equation for the new estimate, $\mathbf{f}^{n+1}$. The convergence of the algorithm depends on the form of the prior energy function, and there is no proof of the convergence of the algorithm in general.

Lange [26] derived several ϕ functions with desirable mathematical properties. These ϕ functions, however, do not promote formation of discontinuities, nor was there empirical testing of the functions. The OSL algorithm was also modified in such a way that searching was done in a particular direction to guarantee convergence.

Lalush and Tsui [25] compared three different ϕ functions proposed in [16, 19, 20] by parameterizing ϕ. They showed that the effects of these ϕ functions could be understood by examining the derivatives of the functions. The motivation here was to run empirical tests on image reconstruction under different parameter settings. In all instances, they used the MAP-EM OSL procedures for reconstruction. The results showed that the ϕ functions proposed by Geman and McClure [16], and by Hebert and Leahy [20] lead to much more selective smoothing than does the ϕ function suggested by Green [19], which tends to smooth noise and edges equally.

5.5.2 Prior Defined on Intensities and Line Processes

Some researchers modeled priors using both line process and intensities to capture the property of discontinuities. The main drawback for using line processes directly, however, is that it leads to a difficult mixed variable optimization problem since there is no objective function, ϕ. Thus the posterior distribution is maximized directly instead of minimizing an objective function.

Johnson, *et al.* [21] modeled the prior using line processes with complex cliques. They mentioned the difficulty of optimization due to the mixed variables. In this study, the Poisson likelihood is approximated by a Gaussian distribution, using the square root of the intensities as variables. Since in a Gaussian distribution, the mode and the mean are the same, they used the Iterative Conditional Average (ICA) method for the intensity updating procedure which is equivalent to performing Iterative Conditional Mode (ICM) updating. The Gaussian form is, however, only an approximation of the true Poisson likelihood. In the Poisson case, ICA and ICM lead to different values.

Since they do not have a ϕ function, they estimate the line processes directly. However, due to the binary nature of line processes, the line estimates will converge faster than the image intensities when ICM is used for both line process update and intensity update. To avoid this, they let the line processes take continuous values in the interval $[0, 1]$ and then use ICA on line processes. The expected values of binary valued line processes, however, are continuous, hence it may have been unnecessary to make the line processes be continuous first and then apply the ICA method. They could have directly used the expected values of line processes from the binary values. Using the continuous line values to evaluate the mean values probably leads to a smooth approximation of their Gibbs prior. There is, further, no proof of convergence of their ICA method.

The approach of Leahy and Yan [30] is perhaps closest in spirit to our work. Their prior consists of the weak membrane energy function with cliques, and the search method involved an ICM procedure on the intensities. Their update equation for the intensities is similar to Equation 5.34, but it explicitly contains line processes since there is no ϕ function in the model. The ICA method with varying temperature control parameter, T, is used to estimate the line processes. The ICA method does not guarantee that the estimated line process converges to zero or one, and the method is not proven to converge to a local minimum. To overcome the above problem, an annealing-like technique was used for the line process estimation within the GEM procedure, and the prior energy function was modified by introducing the control parameter, T. When fixing $T = 1$, their line process update is equivalent to the ICA method. In the limit $T \rightarrow 1$, the estimation becomes equivalent to the ICM method. At each GEM iteration, after updating the intensities, the line processes are updated in an annealing fashion by reducing temperature, although there is no detailed report on how the temperature reduction is done. Reducing the temperature to zero makes the line processes converge to the values zero or one. Their approach is different than ours in that they used the temperature varying scheme with the GEM procedure to estimate line processes. Our approach, on the other hand, employs the temperature scheme globally, and each GEM iteration is carried out at a given temperature. The convergence of Leahy and Yan's algorithm is not guaranteed.

5.6 Results

We simulated the algorithm using the math phantom shown in Figure 5.5(a). The 40x40 phantom contains three "hot" regions of intensity 110 counts, three cold regions of intensity 80, and a background of intensity 100. We used 40 projection angles over 360° with 40 detector bins at each projection angle. Independent Poisson noise was added to each detector bin. The total counts were approximately 2.6 x 10^6 counts. In our algorithm, $\mathcal{H}$ is the Radon transform. Figure 5.6(b) shows

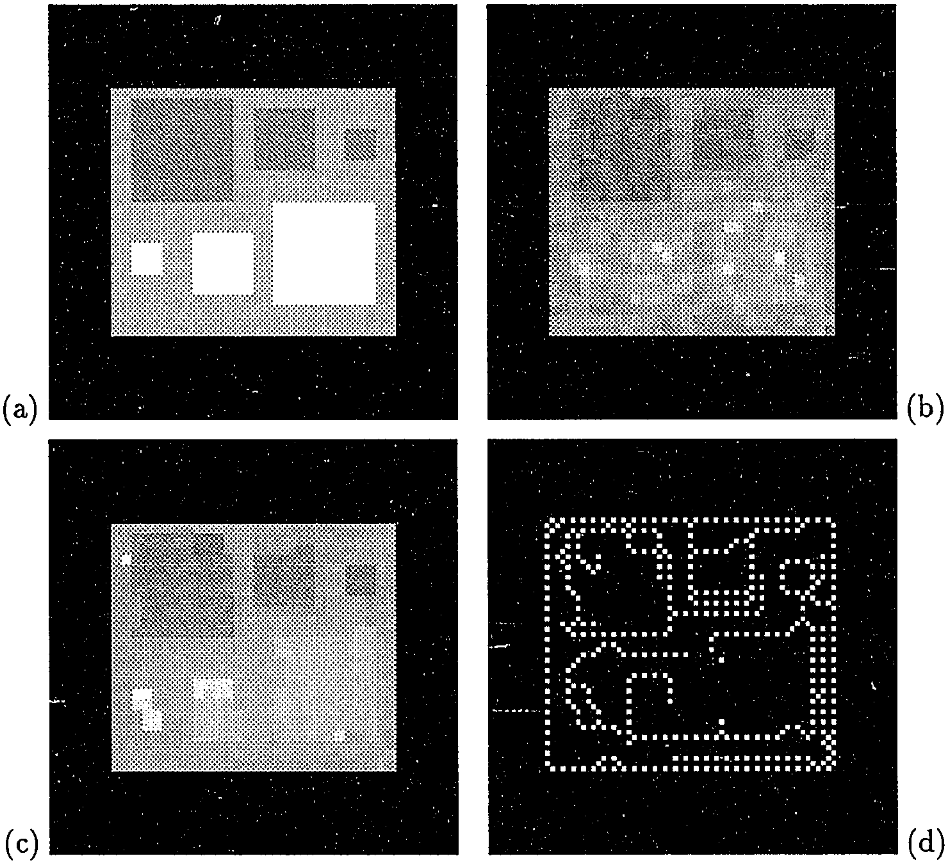


Figure 5.5: Simulations on Phantom IV. (a) 40x40 phantom. (b) EM–ML reconstruction. (c) Reconstruction using quenching algorithm. (d) Expected value of line processes, **z**, in (c).

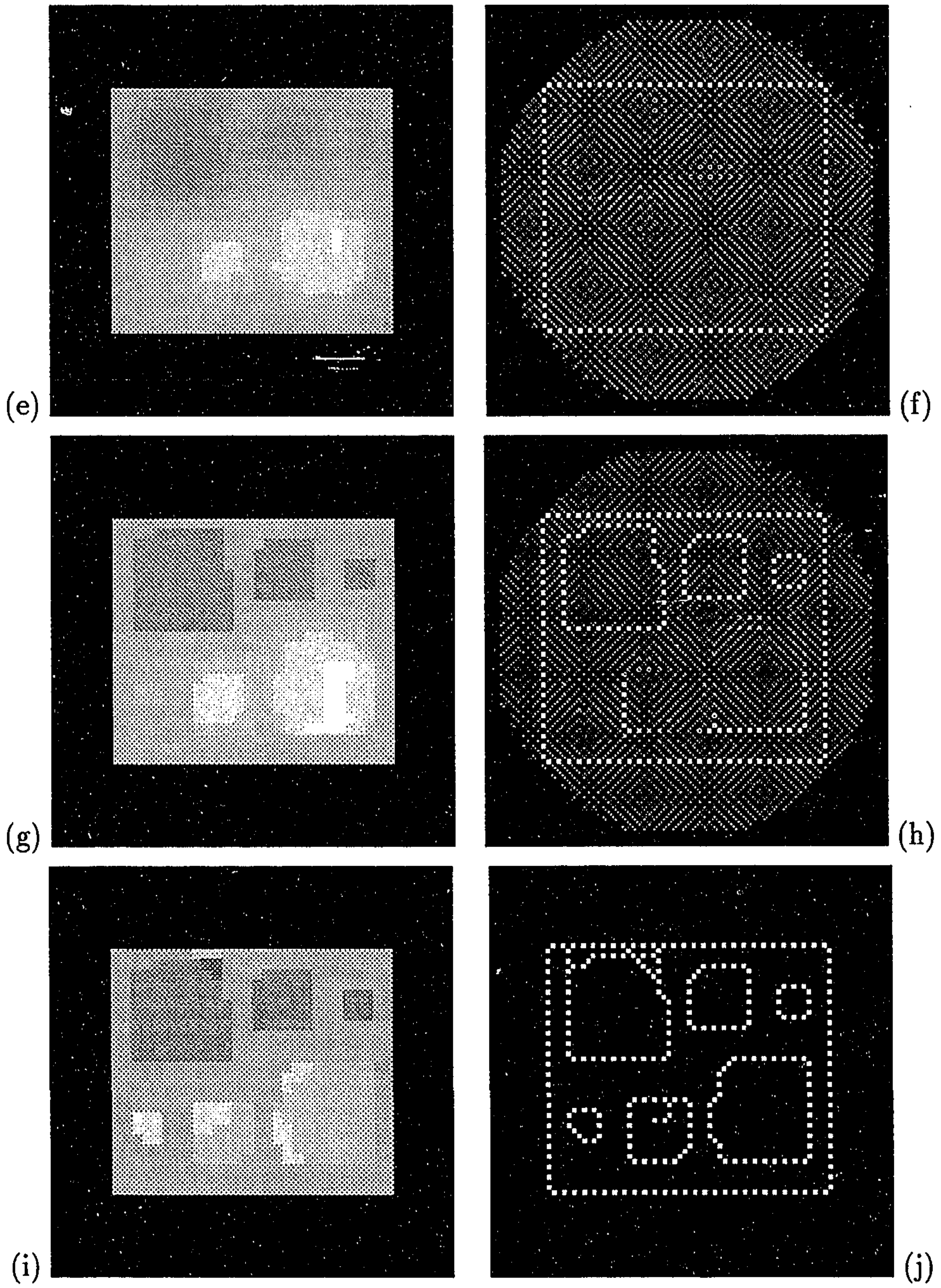


Figure 5.6: Simulations on Phantom IV. (e) Reconstruction using deterministic annealing algorithm at $k = 3$. (f) Expected value of line processes, $\mathbf{z}$ at $k = 3$. (g) and (h) $k = 6$. (i) and (j) $k = 13$.

an EM–ML result obtained after 45 iterations, which corresponds to the minimum RMS error over the entire image. Note that the picture is noisy and that the low contrast hot regions are barely discernible. Figures 5.6(e) – (j) show the source, $\mathbf{f}$, reconstructions and corresponding expected value of the line processes, $\mathbf{z}$, for the deterministic annealing algorithm at high to low temperatures ($k = 3, 6, 13$). Note that all three hot regions are recovered and the corresponding line processes are approximately correct. The initial condition for $\mathbf{f}$ was a constant intensity of 50, and $\mathbf{z}$ was initialized to 0.5. We used 13 values of β with an initial β of 0.03125 and a β doubling annealing schedule. The corresponding temperature at each temperature iteration is shown in Figure 5.7. At a given temperature, we optimize until the objective difference at successive iterations satisfies $|\Delta E| \leq \tau_k$ where k is a temperature iteration number and $\tau_{k+1} = \tau_k/2$ with $\tau_1 = 0.3$. The annealing is terminated when $z_{ij} \leq 0.1$ or $z_{ij} \geq 0.9$, for all (i, j). The parameters were $\lambda_2 = 0.1$ and $\alpha = 2.7$.

We eliminated the annealing strategy and instead ran a "quenching" algorithm by executing the deterministic annealing algorithm at a single, very high value of β ($\beta = 256$) with the same parameters. The results in Figures 5.5(c) and (d) show an inferior reconstruction with profuse and erroneous line processes compared to the annealing results in Figures 5.6(e) – (j).

The computation time for one iteration of our GEM algorithm is close to that of one iteration of a standard EM algorithm. Even though the inclusion of the mean of the line processes triples the number of variables relative to a standard EM update, the line process iterations are computationally negligible compared to

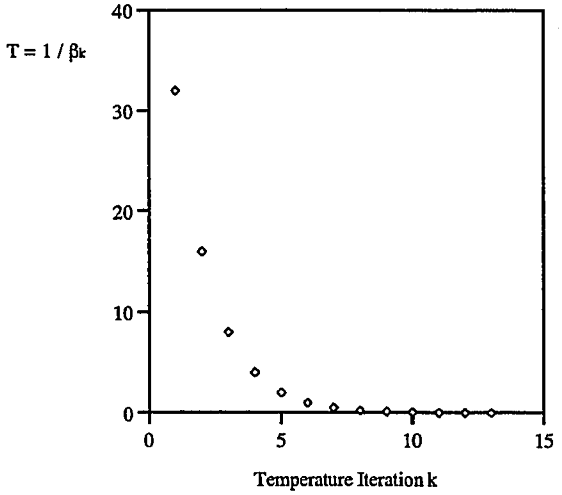


Figure 5.7: Temperature $(1/\beta)$ vs. temperature iteration, k.

the update of the source intensity variables, $\mathbf{f}$. The update of $\mathbf{f}$ is computationally intensive since it involves computation of projections. For our deterministic annealing runs, the number of iterations at a given temperature depends on the choice of threshold τ_1. For $\tau_1 = 0.3$, the number of iterations ranged between 6 and 68 for a total of 311 iterations over 13 temperatures. The EM result was obtained at 45 iterations and the quenched annealing at 60 iterations. To a great extent the computational burden depends on the termination criteria chosen.

Table 5.2 show the quantitative results in the form of RMS errors in the seven regions of interest (ROI) corresponding to the six squares and background of the phantom. The RMS error for the EM algorithm is taken at the 45^{th} iteration. As seen, the best total RMS error was achieved by the deterministic annealing algorithm. Qualitative and quantitative comparison to the quenching method shows

Table 5.2: RMS error in ROI for various algorithms

	EM-ML	Quenching	Deterministic Annealing
Top left	4.935	4.074	4.984
Top middle	5.385	5.271	3.394
Top right	4.831	1.120	0.382
Bottom left	7.046	6.826	6.400
Bottom middle	5.039	3.168	3.449
Bottom right	5.210	1.816	2.314
Base region	5.656	3.183	2.004
Total image	4.293	2.633	2.264

that temperature annealing is beneficial. The RMS errors in the entire image are also plotted in Figure 5.8. Although the total RMS errors for the deterministic annealing algorithm and the quenching algorithm are fairly close, the reconstructed images look much different. The profile along a horizontal line through the upper three squares is shown in Figure 5.9, and through the lower three squares in Figure 5.10. The profile of the deterministic annealing reconstruction is clearly closer to the phantom than the profile of the quenching reconstruction.

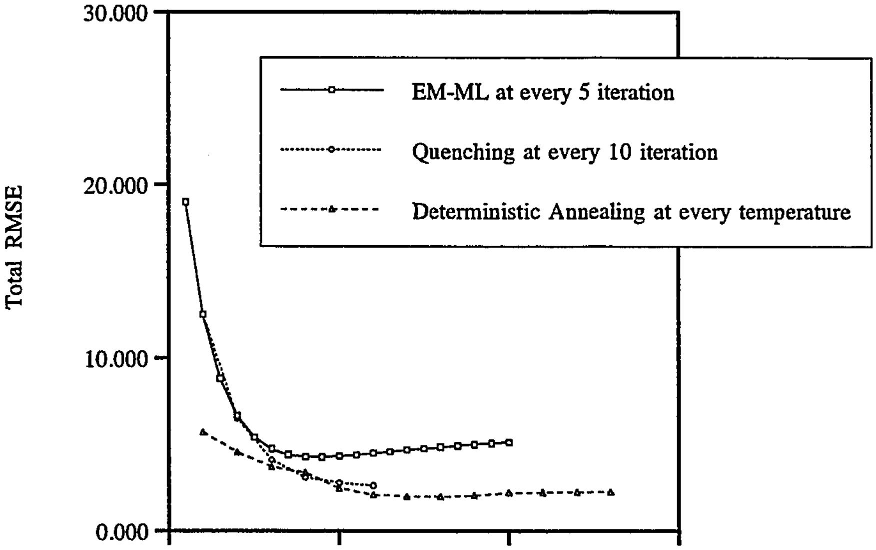


Figure 5.8: Total RMSE for different algorithms used

5.7 Discussion

All related previous efforts have used smooth Gibbs energy functions (or equivalent) to approximate their prior model. It is known that smooth functions are easier to handle. In general, if a potential function has desirable mathematical properties, such as convexity and existence of the first and the second derivatives, it does not promote the formation of discontinuities. On the other hand, if a prior captures the property of discontinuities, either by introducing line processes directly or by using a certain form of ϕ function, it tends to lead to a difficult optimization problem.

One of the advantages of the deterministic annealing approach presented here is that it allows us to approach the broken parabola, which is in fact the correct

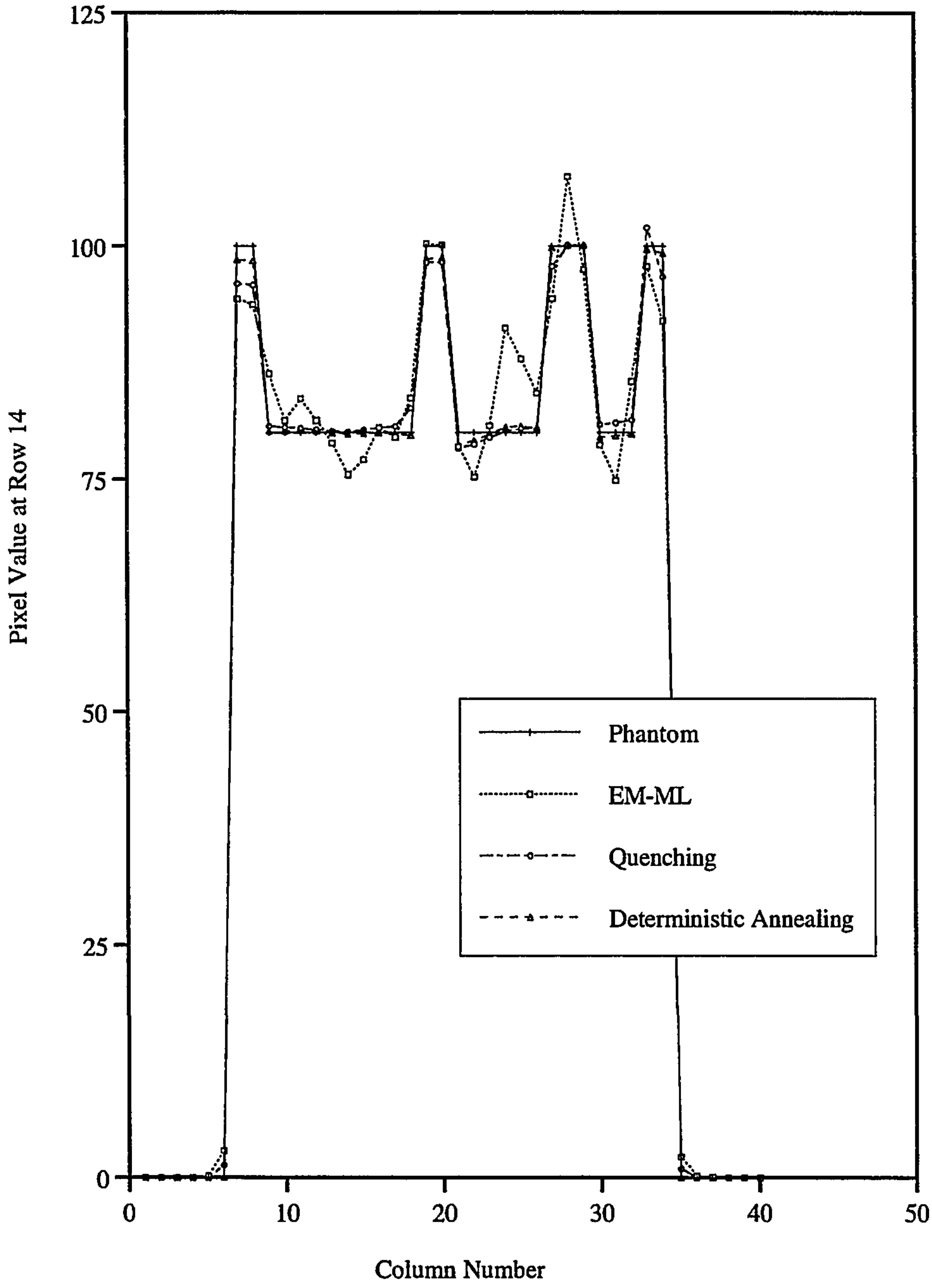


Figure 5.9: Profile along a horizontal line (row 14 from the top).

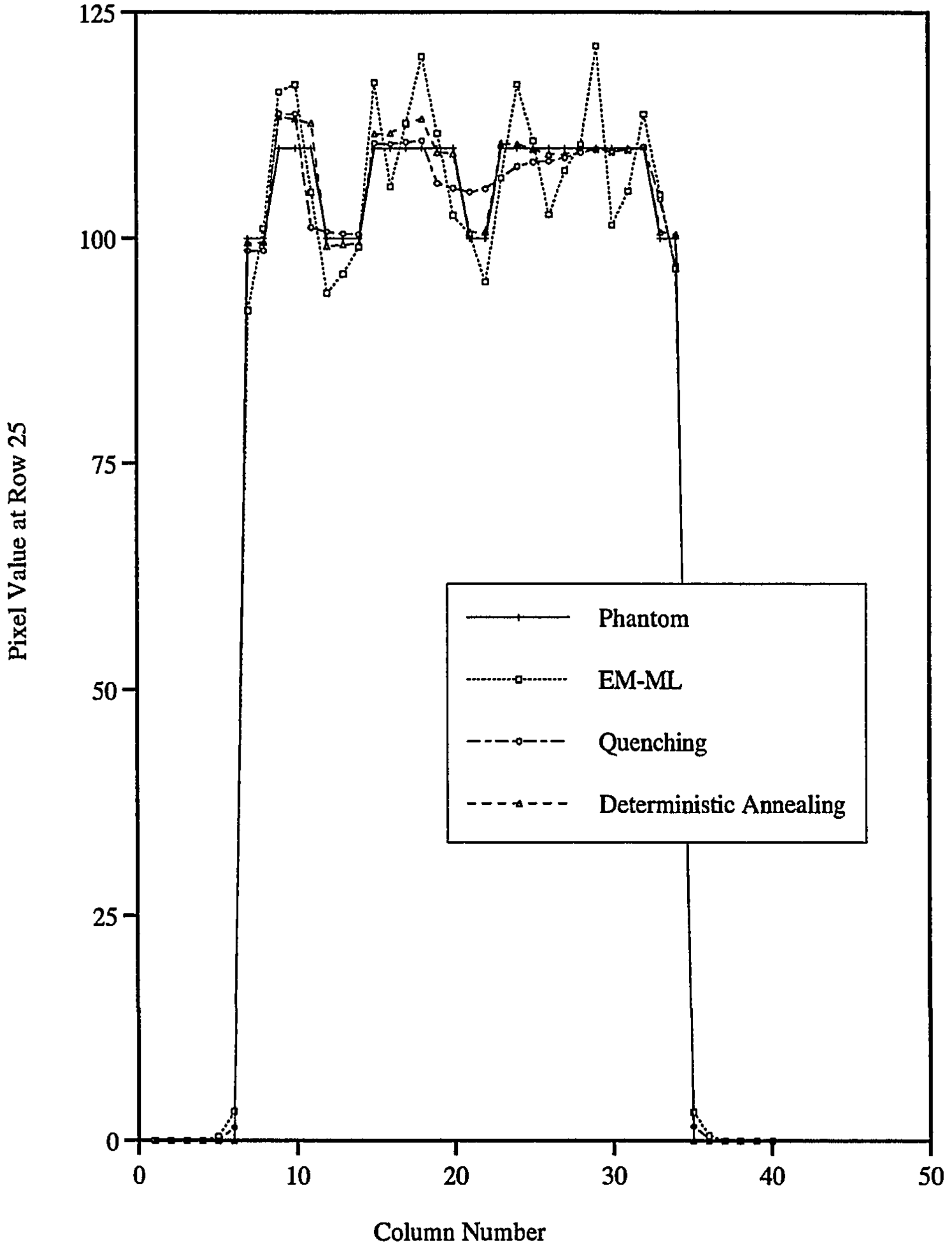


Figure 5.10: Profile along a horizontal line (row 25 from the top).

potential function derived from the original weak membrane energy function that promotes formation of discontinuities, through the variation of a control parameter. Since the correct potential function is approached rather than being approximated, the original intent of the objective function is retained.

The second advantage is that the method supports an efficient search of global minima in the reconstruction problem. The local minimum problem has been addressed by the method of simulated annealing. In essence, simulated annealing consists of sampling from a posterior distribution combined with progressive lowering of the temperature (annealing). It has been shown that there exists [15] an optimal annealing schedule that can be used to find the global minimum. Unfortunately, following this annealing schedule, or even an approximation thereof, is computationally prohibitive rendering simulated annealing impractical. An advantage of the deterministic annealing method is that it avoids stochastic sampling and may handle local minima in an efficient manner. The deterministic algorithm tracks the minimum through a sequence of objective functions that more closely approximate the original objective function as the temperature is lowered. Since the first member of the objective function sequence is a smooth approximation of the original objective, its minimum, which we expect to be in the vicinity of the desired global minimum, is most likely found at a low temperature. The deterministic annealing algorithm is also parallelizable, making it particularly appealing in a parallel computing environment.

The use of the weak membrane model in our work represents a particular choice of prior. The deterministic annealing framework does not depend on this choice and may be extended to more complex situations involving cliques as shown in [13, 45]. As noted earlier, the intermodality information can also be incorporated in ϕ_β by making the weak membrane parameter, α, space variant following Equation 4.15.

As discussed in Section 5.5, there are, in general, tradeoffs between having a mathematically desirable potential function and potential functions that smooth noise and discontinuities selectively. Furthermore, in cases with a prior using line processes to capture the property of discontinuity formation, we end up with a difficult optimization step leading to a non-closed form update procedure. The deterministic annealing method proposed here, which results in closed form update equations and is mathematically more tractable, is one way to overcome some of these obstacles.

Chapter 6

Summary

We have presented two approaches to improving image reconstruction in emission tomography in this thesis. In Chapter 4, we proposed a method that incorporated anatomical information in the reconstruction procedure. In previous efforts, the anatomical information was incorporated as part of the optimization procedure. Here, the anatomical information is incorporated as a part of the Gibbs prior energy function and, therefore, does not depend on a particular algorithm used for the actual optimization. The results show that the reconstructions with intermodality information are improved, not only over EM-ML reconstructions but also over MAP reconstructions that use only a generic smoothness prior. In Chapter 5, we proposed a new framework for emission tomography reconstruction using a continuation method. Although continuation methods have been successfully used in other areas, such as computer vision, they have not yet been applied in the domain of image reconstruction from *projection* data. This is mainly because each projection data measurement is globally connected to many source locations which makes it difficult to employ a continuation method directly. We derived a deterministic annealing algorithm with closed form update equations.

One potential improvement for both of these approaches could be a more systematic way of estimating and/or adjusting parameters such as the prior weighting parameter, λ_1 and λ_2, the intermodality parameters, κ_1 and κ_2, in Chapter 4, and the temperature control parameter, β, in Chapter 5. Geman and McClure [16] used an EM method to estimate the prior weighting parameter λ. Reconstructions under different prior weighting parameter settings have been investigated in [25]. The intermodality parameters can perhaps capture particular edge information. For instance, we know that the typical radiopharmaceutical uptake ratio in grey and white matter of the brain is 4:1. This ratio may be used to estimate κ_1 and κ_2 for grey and white matter edges. As discussed in Section 4.5, further validation studies of the correlation between anatomical structures and functions for different types of radiopharmaceutical are necessary for incorporating anatomical information correctly. Development of a sophisticated segmentation algorithm could also help us to select only valid anatomical edges.

Future directions include incorporation of intermodality information in the deterministic annealing algorithm and modeling attenuation corrections in the forward projection operator $\mathcal{H}$ to enable our algorithms to be applied to clinical data. The framework presented in Chapter 5 could also be used with other priors. Work is being done using this framework with the *weak plate* prior [32], which allows piecewise linear regions in the reconstruction rather than piecewise flat regions enforced by a weak membrane prior.

Appendix A

In this appendix, we prove the relation Equation 3.15 regarding the conditional expectation, namely,

$$E_{\mathrm{C}}\{C_{t\theta;ij}|\mathbf{G}=\mathbf{g},\hat{\mathbf{f}}^n\} = g_{t\theta}\frac{\mathcal{H}_{t\theta;ij}\hat{f}_{ij}^n}{\sum_{kl}\mathcal{H}_{t\theta;kl}\hat{f}_{kl}^n}.$$

Let us consider k number of independent Poisson random variables $X_1, \cdots, X_k$ with means $\mu_1, \cdots, \mu_k$. The joint probability is then

$$\Pr(X_1 = x_1, \cdots, X_k = x_k) = e^{-(\mu_1+\cdots+\mu_k)}\frac{{\mu_1}^{x_1}}{x_1!}\cdots\frac{{\mu_k}^{x_k}}{x_k!}.$$

It it also known that the sum of independent Poisson random variables is a Poisson random variable [46]. We, therefore, can write:

$$\Pr(Y = y) \;=\; e^{-\mu}\frac{\mu^y}{y!}$$

where $Y = X_1 + \cdots + X_k$ and $\mu = \mu_1 + \cdots + \mu_k$. Then the joint probability of $X_1 = x_1, \cdots, X_k = x_k$ conditioned on their sum $Y = y$,

$$\begin{aligned} \Pr(X_1 = x_1, \cdots, X_k = x_k | Y = y) &= \frac{\Pr(X_1 = x_1, \cdots, X_k = x_k)}{\Pr(Y = y)} \\ &= \frac{y!}{x_1! \cdots x_k!} \left(\frac{\mu_1}{\mu}\right)^{x_1} \cdots \left(\frac{\mu_k}{\mu}\right)^{x_k}. \end{aligned}$$

This is a multinomial distribution of k classes with probabilities $\mu_1/\mu, \cdots, \mu_k/\mu$. The expected value of X_i conditioned on $\sum_j X_j = Y = y$ is

$$\begin{aligned} E\{X_i | Y = y\} &= y\frac{\mu_i}{\mu} \\ &= y\frac{\mu_i}{\sum_j \mu_j}. \end{aligned}$$

Following the derivation found in [20], we consider the complete data, $C_{t\theta;ij}$, which are modeled as independent Poisson random variables with means $\mathcal{H}_{t\theta;ij} f_{ij}$. The joint probability given $\hat{\mathbf{f}}^n$ is then

$$\Pr(\mathbf{C} = \mathbf{c} | \hat{\mathbf{f}}^n) = \prod_{t\theta} \prod_{ij} e^{-\mathcal{H}_{t\theta;ij} \hat{f}_{ij}^n} \frac{\left(\mathcal{H}_{t\theta;ij} \hat{f}_{ij}^n\right)^{c_{t\theta;ij}}}{c_{t\theta;ij}!}.$$

Recall the Poisson likelihood, $\Pr(\mathbf{G} = \mathbf{g} | \mathbf{f})$, in Equation 3.8. The joint probability given $\hat{\mathbf{f}}^n$ can be written

$$\Pr(\mathbf{G} = \mathbf{g} | \hat{\mathbf{f}}^n) = \prod_{t\theta} \frac{\left(\sum_{ij} \mathcal{H}_{t\theta;ij} \hat{f}_{ij}^n\right)^{g_{t\theta}} e^{-\sum_{ij} \mathcal{H}_{t\theta;ij} \hat{f}_{ij}^n}}{g_{t\theta}!}.$$

Since $\sum_{ij} C_{t\theta;ij} = G_{t\theta}$, we have

$$\begin{aligned}&\Pr(C_{t\theta;11} = c_{t\theta;11}, \cdot\cdot, C_{t\theta;ij} = c_{t\theta;ij}, \cdot\cdot, C_{t\theta;MN} = c_{t\theta;MN} | \mathbf{G} = \mathbf{g}, \mathbf{f}^n)\\ &\quad = \Pr(C_{t\theta;11} = c_{t\theta;11}, \cdot\cdot, C_{t\theta;ij} = c_{t\theta;ij}, \cdot\cdot, C_{t\theta;MN} = c_{t\theta;MN} | G_{t\theta} = g_{t\theta}, \mathbf{f}^n),\end{aligned}$$

and

$$E_{\mathbf{C}}\{C_{t\theta;ij} | \mathbf{G} = \mathbf{g}, \mathbf{f}^n\} = E_{\mathbf{C}}\{C_{t\theta;ij} | G_{t\theta} = g_{t\theta}, \mathbf{f}^n\}.$$

The joint distribution of $C_{t\theta;11} = c_{t\theta;11}, \cdots, C_{t\theta;MN} = c_{t\theta;MN}$ for MN independent random variables conditioned on their sum $G_{t\theta} = g_{t\theta}$,

$$\begin{aligned}&\Pr(C_{t\theta;11} = c_{t\theta;11}, \cdot\cdot, C_{t\theta;ij} = c_{t\theta;ij}, \cdot\cdot, C_{t\theta;MN} = c_{t\theta;MN} | G_{t\theta} = g_{t\theta}, \mathbf{f}^n)\\ &= \frac{g_{t\theta}!}{c_{t\theta;11}! \cdots c_{t\theta;MN}!} \left(\frac{\mathcal{H}_{t\theta;11} \hat{f}^n_{11}}{\sum_{kl} \mathcal{H}_{t\theta;kl} \hat{f}^n_{kl}} \right)^{c_{t\theta;11}} \cdots \left(\frac{\mathcal{H}_{t\theta;MN} \hat{f}^n_{MN}}{\sum_{kl} \mathcal{H}_{t\theta;kl} \hat{f}^n_{kl}} \right)^{c_{t\theta;MN}},\end{aligned}$$

is a multinomial distribution of MN classes with probabilities $\frac{\mathcal{H}_{t\theta;11} \hat{f}^n_{11}}{\sum_{kl} \mathcal{H}_{t\theta;kl} \hat{f}^n_{kl}}, \cdots, \frac{\mathcal{H}_{t\theta;MN} \hat{f}^N_{MN}}{\sum_{kl} \mathcal{H}_{t\theta;kl} \hat{f}^n_{kl}}$. Thus the expected value of a multinomial variable $E_{\mathbf{C}}\{C_{t\theta;ij} | G_{t\theta} = g_{t\theta}, \mathbf{f}^n\}$, is equal to

$$g_{t\theta} \frac{\mathcal{H}_{t\theta;ij} \hat{f}^n_{ij}}{\sum_{kl} \mathcal{H}_{t\theta;kl} \hat{f}^n_{kl}}.$$

Appendix B

Algebraic Transformation of the Poisson Likelihood Function

Here, we apply an algebraic transformation to the Poisson log-likelihood objective function and show that a coordinate-wise optimization on the transformation leads to the EM-ML algorithm in Chapter 3. Recall the Poisson log-likelihood function to be maximized (in Equation 3.10),

$$\sum_{t\theta} g_{t\theta} \log(\sum_{ij} \mathcal{H}_{t\theta;ij} f_{ij}) - \sum_{t\theta} \sum_{ij} \mathcal{H}_{t\theta;ij} f_{ij}. \tag{B.1}$$

The work in this section is done is close conjunction with the work of Anand Rangarajan

Consider the following transformation to be applied at every location (i,j):

$$
\begin{aligned}
& \sigma_{t\theta;ij} = \log(\mathcal{H}_{t\theta;ij} f_{ij}) \\
\Rightarrow\; & \mathcal{H}_{t\theta;ij} f_{ij} = e^{\sigma_{t\theta;ij}} \\
\Rightarrow\; & \sum_{ij} \mathcal{H}_{t\theta;ij} f_{ij} = \sum_{ij} e^{\sigma_{t\theta;ij}} \\
\Rightarrow\; & \log(\sum_{ij} \mathcal{H}_{t\theta;ij} f_{ij}) = \log(\sum_{ij} e^{\sigma_{t\theta;ij}}). \qquad \text{(B.2)}
\end{aligned}
$$

We rewrite the log-likelihood function (Equation B.1) using the transform in Equation B.2:

$$
E_{LL} = \sum_{t\theta} g_{t\theta} \log(\sum_{ij} e^{\sigma_{t\theta;ij}}) + \sum_{t\theta} \sum_{ij} u_{t\theta;ij} (\log \mathcal{H}_{t\theta;ij} f_{ij} - \sigma_{t\theta;ij}) - \sum_{t\theta} \sum_{ij} \mathcal{H}_{t\theta;ij} f_{ij}, \qquad \text{(B.3)}
$$

where $u_{t\theta;ij}$ is a Lagrange parameter to be determined. Note that the transformed objective function, E_{LL}, and the original Poisson log-likelihood function have the same fixed points [38].

Now, taking the partial derivatives of E_{LL} with respect to $\sigma_{t\theta;ij}$ and to $u_{t\theta;ij}$ and setting each equal to zero yields,

$$
\begin{aligned}
\frac{\partial E_{LL}}{\partial \sigma_{t\theta;ij}} = 0 \;\Rightarrow\; & \frac{g_{t\theta} e^{\sigma_{t\theta;ij}}}{\sum_{kl} e^{\sigma_{t\theta;kl}}} - u_{t\theta;ij} = 0 \\
\Rightarrow\; & u_{t\theta;ij} = \frac{g_{t\theta} e^{\sigma_{t\theta;ij}}}{\sum_{kl} e^{\sigma_{t\theta;kl}}} \qquad \text{(B.4)} \\
\frac{\partial E_{LL}}{\partial u_{t\theta;ij}} = 0 \;\Rightarrow\; & \sigma_{t\theta;ij} = \log \mathcal{H}_{t\theta;ij} f_{ij}. \qquad \text{(B.5)}
\end{aligned}
$$

Notice that the Lagrange transformation (Equation B.4) and Equation B.5 together form the expected value, $E_{\mathbf{C}}\{C_{t\theta;ij}|\mathbf{G} = \mathbf{g}, \mathbf{f}^n\}$, the E-step (Equation 3.15) of the EM algorithm. The objective function after Lagrange parameter elimination is

$$\begin{aligned} E_{LL} &= \sum_{t\theta} g_{t\theta} \log(\sum_{ij} e^{\sigma_{t\theta;ij}}) + \sum_{t\theta}\sum_{ij} g_{t\theta} \frac{e^{\sigma_{t\theta;ij}}}{\sum_{kl} e^{\sigma_{t\theta;kl}}} \log \mathcal{H}_{t\theta;ij} f_{ij} \\ &\quad - \sum_{t\theta}\sum_{ij} \sigma_{t\theta;ij} g_{t\theta} \frac{e^{\sigma_{t\theta;ij}}}{\sum_{kl} e^{\sigma_{t\theta;kl}}} - \sum_{t\theta}\sum_{ij} \mathcal{H}_{t\theta;ij} f_{ij}. \end{aligned} \tag{B.6}$$

Now, we perform the coordinate-wise optimization on E_{LL} with respect to both $\sigma_{t\theta;ij}$ and f_{ij}. Taking a partial derivative of E_{LL} with respect to $\sigma_{t\theta;ij}$ yields,

$$\frac{\partial E_{LL}}{\partial \sigma_{t\theta;ij}} = g_{t\theta} \left[\frac{e^{\sigma_{t\theta;ij}}}{\sum_{kl} e^{\sigma_{t\theta;kl}}} - \left(\frac{e^{\sigma_{t\theta;ij}}}{\sum_{kl} e^{\sigma_{t\theta;kl}}} \right)^2 \right] (\log \mathcal{H}_{t\theta;ij} - \sigma_{t\theta;ij}) = 0 \tag{B.7}$$

$$\Rightarrow \quad \sigma_{t\theta;ij} = \log \mathcal{H}_{t\theta;ij} f_{ij}. \tag{B.8}$$

Taking a partial derivative of E_{LL} with respect to f_{ij} yields,

$$\frac{\partial E_{LL}}{\partial f_{ij}} = 0 \quad \Rightarrow \quad \sum_{t\theta} g_{t\theta} \frac{e^{\sigma_{t\theta;ij}}}{\sum_{kl} e^{\sigma_{t\theta;kl}}} \frac{1}{f_{t\theta}} - \sum_{t\theta} \mathcal{H}_{t\theta;ij} = 0 \tag{B.9}$$

$$\Rightarrow \quad f_{ij} = \frac{\sum_{t\theta} g_{t\theta;ij} \frac{e^{\sigma_{t\theta;ij}}}{\sum_{kl} e^{\sigma_{t\theta;kl}}}}{\sum_{t\theta} \mathcal{H}_{t\theta;ij}}. \tag{B.10}$$

To ensure that the solution obtained for f_{ij} by taking partial derivative of E_{LL} and setting it to zero is a maximum, it is sufficient to show that the second derivatives evaluated at the fixed point are negative:

$$\left.\frac{\partial^2 E_{LL}}{\partial \sigma^2_{t\theta;ij}}\right|_{\sigma_{t\theta;ij}=\log \mathcal{H}_{t\theta;ij} f_{ij}} = -g_{t\theta}\left.\left[\frac{e^{\sigma_{t\theta;ij}}}{\sum_{kl} e^{\sigma_{t\theta;kl}}} - \left(\frac{e^{\sigma_{t\theta;ij}}}{\sum_{kl} e^{\sigma_{t\theta;kl}}}\right)^2\right]\right|_{\sigma_{t\theta;ij}=\log \mathcal{H}_{t\theta;ij} f_{ij}} \leq 0 \qquad \text{(B.11)}$$

$$\left.\frac{\partial^2 E_{LL}}{\partial f^2_{ij}}\right|_{\sigma_{t\theta;ij}=\log \mathcal{H}_{t\theta;ij} f_{ij}} = -\sum_{t\theta} g_{t\theta}\frac{e^{\sigma_{t\theta;ij}}}{\sum_{kl} e^{\sigma_{t\theta;kl}}}\frac{1}{f^2_{ij}} \leq 0. \qquad \text{(B.12)}$$

In practice, we do not let f_{ij} vanish completely. Therefore, the cases where the second derivatives in Equation B.11 and Equation B.12 go to zero do not occur and the fixed point is indeed a maximum. The general convergence of the coordinate-wise optimization to a fixed point is shown in [36].

The coordinate-wise update equations can be written

$$\hat{f}^{n+1}_{ij} = \frac{\sum_{t\theta} g_{t\theta;ij}\frac{e^{\hat{\sigma}^n_{t\theta;ij}}}{\sum_{kl} e^{\hat{\sigma}^n_{t\theta;kl}}}}{\sum_{t\theta} \mathcal{H}_{t\theta;ij}}$$

$$\hat{\sigma}^{n+1}_{t\theta;kl} = \log \mathcal{H} \hat{f}^{n+1}_{ij}.$$

The above two equations can be combined into one update equation for f_{ij}:

$$\hat{f}^{n+1}_{ij} = \frac{\sum_{t\theta} g_{t\theta;ij}\frac{\mathcal{H}\hat{f}^n_{ij}}{\sum_{kl} \mathcal{H}\hat{f}^n_{ij}}}{\sum_{t\theta} \mathcal{H}_{t\theta;ij}}. \qquad \text{(B.13)}$$

Notice that the above update equation is the same as the EM–ML update equation for f_{ij} in Equation 3.18.

We can also obtain the M-step function of the EM-ML method in Equation 3.16 by substituting $\log \mathcal{H}_{t\theta;ij} \hat{f}^n_{ij}$ for $\sigma_{t\theta;ij}$ in Equation B.6:

$$\begin{aligned} E_{LL} &= \sum_{t\theta} g_{t\theta} \log\Big(\sum_{ij} \mathcal{H}_{t\theta;ij} \hat{f}^n_{ij}\Big) + \sum_{t\theta} \sum_{ij} g_{t\theta} \frac{\mathcal{H}_{t\theta;ij} \hat{f}^n_{ij}}{\sum_{kl} \mathcal{H}_{t\theta;ij} \hat{f}^n_{kl}} \log \mathcal{H}_{t\theta;ij} f_{ij} \\ &\quad - \sum_{t\theta} \sum_{ij} g_{t\theta} \log(\mathcal{H}_{t\theta;ij} \hat{f}^n_{ij}) \frac{\mathcal{H}_{t\theta;ij} \hat{f}^n_{ij}}{\sum_{kl} \mathcal{H}_{t\theta;ij} \hat{f}^n_{kl}} - \sum_{t\theta} \sum_{ij} \mathcal{H}_{t\theta;ij} f_{ij}. \\ &= \sum_{t\theta} \sum_{ij} g_{t\theta} \frac{\mathcal{H}_{t\theta;ij} \hat{f}^n_{ij}}{\sum_{kl} \mathcal{H}_{t\theta;kl} \hat{f}^n_{kl}} \log f_{ij} - \sum_{t\theta} \sum_{ij} \mathcal{H}_{t\theta;ij} f_{ij} \\ &\quad + \text{terms independent of } \mathbf{f}. \end{aligned} \tag{B.14}$$

We have shown that a coordinate-wise optimization on the algebraically transformed function of the Poisson log-likelihood objective function is equivalent to the EM–ML method in Section 3.2.1.1, and that it converges to a local maximum.

Bibliography

[1] N. C. Andreasen. "Brain Imaging: Applications in Psychiatry". *Science*, 239:1381–1388, 1988.

[2] H. H. Barrett and W. Swindell. *Radiological Imaging: The Theory of Image Formation, Detection, and Processing.* Academic Press, New York, 1981.

[3] J. Besag. "Spatial Interaction and the Statistical Analysis of Lattice Systems (with discussion)". *Journal of the Royal Statistical Society, B*, 36:192–326, 1986.

[4] M. Black and A. Rangarajan. "On Line Processes, Outlier Rejection, and Robust Statistics". Technical Report YALEU-DCS-RR-993, Department of Computer Science, Yale University, October 1993.

[5] A. Blake and A. Zisserman. *Visual Reconstruction.* Artificial Intelligence. MIT Press, Cambridge, MA, 1987.

[6] L. T. Chang. "A Method for Attenuation Correction in Radionuclide Computed Tomography". *IEEE Trans. on Nuclear Science*, NS-25(1):638–643, 1978.

[7] C. T. Chen, X. Ouyang, W. H. Wong, X. Hu, V. E. Johnson, C. Ordonez, and C. E. Metz. "Sensor Fusion in Image Reconstruction". *IEEE Trans. Nuc. Sci.*, NS-38(2):687–692, April 1991.

[8] C. T. Chen, C. A. Pelizzari, G. T. Y. Chen, M. D. Cooper, and D. N. Levin. "Image Analysis of PET Data with the Aid of CT and MR Images". In C. N. deGraaf and M. A. Viergever, editors, *Information Processing in Medical Imaging*, page 601. Plenum Press, 1987.

[9] B. Y. Croft. *Single-Photon Emission Computed Tomography.* Year Book Medical Publishers, Chicago, 1986.

[10] A. P. Dempster, N. M. Laird, and D. B. Rubin. "Maximum Likelihood from Incomplete Data via the EM Algorithm". *J. R. Statist. Soc.*, 39:1–38, 1977.

[11] P. J. Ell and B. L. Holman, editors. *Computed Emission Tomography.* Oxford University Press, Oxford, 1982.

[12] E. Gamble, D. Geiger, T. Poggio, and D. Weinshall. "Integration of Vision Modules and Labeling of Surface Discontinuities". *IEEE Trans. on Sys., Man and Cybernetics*, SMC-19(6):1576–1581, November/December 1989.

[13] D. Geiger and F. Girosi. "Parallel and Deterministic Algorithms from MRFs: Surface Reconstruction". *IEEE Trans. on Pattern Analysis and Machine Intelligence*, PAMI-13(5):401–412, May 1991.

[14] D. Geman and G. Reynolds. "Constrained Restoration and the Recovery of Discontinuities". *IEEE Trans. on Pattern Analysis and Machine Intelligence*, PAMI-14(3):367–383, March 1992.

[15] S. Geman and D. Geman. "Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images". *IEEE Trans. on Pattern Analysis and Machine Intelligence*, PAMI-6(6):721–741, November 1984.

[16] S. Geman and D. E. McClure. "Statistical Methods for Tomographic Image Reconstruction". In *Proceedings of the 46th Session of the ISI, Bulletin of the ISI*, 1987.

[17] P. Gerlot and Y. Bizais. "Image Registration: A Review and a Strategy for Medical Applications". In C. N. deGraaf and M. A. Viergever, editors, *Information Processing in Medical Imaging*, pages 81–89. Plenum Press, 1987.

[18] G. Gindi, M. Lee, A. Rangarajan, and I. G. Zubal. "Bayesian Reconstruction of Functional Images Using Registered Anatomical Images as Priors". *IEEE Trans. on Medical Imaging*, TMI-12(4):670–680, December 1993.

[19] P. J. Green. "Bayesian Reconstructions from Emission Tomography Data Using A Modified EM Algorithm". *IEEE Trans. on Medical Imaging*, MI-9(1):84–93, March 1990.

[20] T. Hebert and R. Leahy. "A Generalized EM Algorithm for 3-D Bayesian Reconstruction for Poisson Data using Gibbs Priors". *IEEE Trans. on Medical Imaging*, MI-8(2):194–202, June 1989.

[21] V. E. Johnson, W. H. Wong, X. Hu, and C. T. Chen. "Image Restoration Using Gibbs Priors: Boundary Modeling, Treatment of Blurring, and Selection of Hyperparameter". *IEEE Trans. Patt. Anal. Mach. Intell.*, PAMI-13(5):413–425, May 1991.

[22] G. Knoll. *Radiation Detection and Measurement.* Wiley, New York, 1989.

[23] D. E. Kuhl. "Rotational Scanning of the Liver". *Radiology*, 71:875–876, 1958.

[24] D. E. Kuhl, J. Engel Jr, M. E. Phelps, and C. Selin. "Epileptic Patterns of Local Cerebral Metabolism and Perfusion in Humans Determined by Emission Computed Tomography of ^{18}FDG and ^{13}NH$_3$". *Annuals of Neurology*, 8(4):348–360, October 1980.

[25] D. S. Lalush and B. M. W. Tsui. "Simulation Evaluation of Gibbs Prior Distributions for Use in Maximum *A Posteriori* SPECT Reconstructions". *IEEE Trans. on Medical Imaging*, MI-11:267–275, 1992.

[26] K. Lange. "Convergence of EM Image Reconstruction Algorithms with Gibbs Smoothing". *IEEE Trans. on Medical Imaging*, MI-9(4):439–446, December 1990.

[27] K. Lange and R. Carson. "EM Reconstruction Algorithm for Emission and Transmission Tomography". *Journal of Computer Assisted Tomography*, 8:306–316, April 1984.

[28] P. C. Lauterbur. "Image Formation by Induced Local Interactions: Examples Employing Nuclear Magnetic Resonance". *Nature*, 242:190–191, 1973.

[29] R. Leahy, T. Hebert, and R. Lee. "Applications of Markov Random Fields in Medical Imaging". In D. A. Ortendahl and J. Llacer, editors, *Information Processing in Medical Imaging*, pages 1–14. Wiley-Liss, 1989.

[30] R. Leahy and X. Yan. "Incorporation of Anatomical MR Data for Improved Functional Imaging with PET". In A. C. F. Colchester and D. J. Hawkes, editors, *Information Processing in Medical Imaging*, pages 105–120. Springer–Verlag, 1991.

[31] M. Lee, A. Rangarajan, I. G. Zubal, and G. Gindi. "A Continuation Method for Emission Tomography". *IEEE Trans. on Nuclear Science*, 40(6):2049–2058, 1993.

[32] S.-J. Lee, A. Rangarajan, and G. Gindi. "Weak Plate Mechanical Models in Bayesian Reconstruction for Emission Tomography". In *Conference Record of the 1993 IEEE Nuclear Science Symposium and Medical Imaging Conference (in press)*.

[33] E. Levitan and G. T. Herman. "A Maximum *A Posteriori* Probability Expectation Maximization Algorithm for Image Reconstruction in Emission Tomography". *IEEE Trans. on Medical Imaging*, MI-6:185–192, 1987.

[34] Z. Liang, R. Jasczack, R. Coleman, and V. Johnson. "Simultaneous Reconstruction, Segmentation and Edge Enhancement of Relatively Piecewise Continuous Images with Intensity Level Information". *Medical Physics*, 18(3):394–401, May/June 1991.

[35] J. Llacer and E. Veklerov. "Feasible Images and Practical Stopping Rules for Iterative Algorithms in Emission Tomography". *IEEE Trans. on Medical Imaging*, 8, 1989.

[36] D. G. Luenberger. *Linear and Nonlinear Programming.* Addision-Wesley Publishing Company, Reading, MA, second edition, 1984.

[37] P. Mansfield and P. K. Grannell. "NMR Diffraction in Solids". *Journal of Phys C: Solid State Physics*, 6:L422–426, 1973.

[38] E. Mjolsness and C. Garrett. "Algebraic Transformations of Objective Functions". *Neural Networks*, 3:651–669, 1990.

[39] P. G. Morris. *Nuclear Magnetic Resonance Imaging in Medicine and Biology.* Clarendon Press, Oxford, 1986.

[40] K. Nakajima, N. Shuke, J. Taki, T. Ichihara, N. Motomura, H. Bunko, and K. Hisada. "A Simulation of Dynamic SPECT Using Radiopharmaceuticals with Rapid Clearance". *The Journal of Nuclear Medicine*, 33(6):1200–1206, 1992.

[41] M. E. Phelps, E. J. Hoffman, N. A. Mullani, and M. M. Ter-Pogossian. "Application of Annihilation Coincidence Detection to Transaxial Reconstruction Tomography". *Journal of Nuclear Medicine*, 16:210, 1975.

[42] M. E. Phelps, J. C. Mazziotta, and H. R. Schelbert. *Positron Emission Tomography and Autoradiography: Principles and Applications for the Brain and Heart.* Raven Press, New York, 1986.

[43] T. Poggio, E. B. Gamble, and J. J. Little. "Parallel Integration of Vision Modules". *Science*, 242:436–440, October 1988.

[44] H. V. Poor. *An Introduction to Signal Detection and Estimation.* Springer-Verlag, New York, 1988.

[45] A. Rangarajan and R. Chellappa. "A Continuation Method for Image Estimation and Segmentation". Technical Report CAR-TR-586, Center for Automation Research, University of Maryland, October 1991.

[46] C. R. Rao. *Advanced Statistical Methods in Biometric Research.* Hafner Publishing Company, Darien, Conn, 1970.

[47] C. R. Rao. *Linear Statistical Inference and Its Applications.* Wiley, New York, 1973.

[48] A. Rosenfeld and A. C. Kak, editors. *Digital Picture Processing: Volume 1.* Academic Press, New York, second edition, 1982.

[49] E. Serge. *Nuclei and Particles: An Introduction to Nuclear and Subnuclear Physics.* W. A. Benjamin, Inc., New York, 1965.

[50] L. A. Shepp and Y. Vardi. "Maximum Likelihood Reconstruction for Emission Tomography". *IEEE Trans. on Medical Imaging*, MI-1(2):113–122, October 1982.

[51] D. L. Snyder and M. I. Miller. "The Use of Sieves to Stabilize Images Produced with the EM Algorithm for Emission Tomography". *IEEE Trans. on Nuclear Science*, NS-32:3864–3872, 1985.

[52] D. L. Snyder, M. I. Miller, L. J. Thomas Jr., and D. G. Politte. "Noise and Edge Artifacts in Maximum-Likelihood Reconstructions for Emission Tomography". *IEEE Trans. on Medical Imaging*, MI-6(3):228–238, 1987.

[53] J. A. Sorenson and M. E. Phelps. *Physics in Nuclear Medicine.* Grune and Stratton, Orlando, second edition, 1987.

[54] J. R. Thompson and R. A. Tapia. *Nonparametric Function Estimation, Modeling, and Simulation.* Society for Industrial and Applied Mathematics, Philadelphia, 1990.

[55] H. Van Trees. *Detection, Estimation, and Modulation Theory: Part I.* John Wiley and Sons, New York, 1968.

[56] Y. Vardi, L. A. Shepp, and L. Kaufman. "A Statistical Model for Positron Emission Tomography". *J. Amer. Stat. Assoc*, 80:8–37, 1985.

[57] R. C. Walovitch, T. C. Hill, S. T. Garrity, E. H. Cheesman, B. A. Burgess, D. H. O'Leary, A. D. Watson, M. V. Ganey, R. A. Morgan, and S. J. Willams. "Characterization of Technetium-99m-L,L-ECD for Brain Perfusion Imaging, Part 1: Pharmacology of Technetium-99m ECD in Nonhuman Primates". *The Journal of Nuclear Medicine*, 30(11):1892–1901, 1989.

[58] C. F. J. Wu. "On the Convergence Properties of the EM Algorithm". *The Annals of Statistics*, 11(1):95–103, 1983.

[59] G. Zubal and C. Harrell. "Voxel Based Monte Carlo Calculations of Nuclear Medicine Images and Applied Variance Reduction Techniques". In A. C. F.

Colchester and D. J. Hawkes, editors, *Information Processing in Medical Imaging*, pages 23–33. Springer–Verlag, 1991.

[60] I. G. Zubal, G. Gindi, M. Lee, C. Harrell, and E. Smith. "High Resolution Anthropomorphic Phantom for Monte Carlo Analysis of Internal Radiation Sources". In *Proceedings of the Third Annual IEEE Symposium on Computer-Based Medical Systems*, pages 540–547, 1990.

[61] I. G. Zubal, C. R. Harrell, and P. D. Esser. "Monte Carlo Determination of Emerging Energy Spectra for Diagnostically Realistic Radiopharmaceutical Distributions". *Nuclear Instruments and Methods in Physics Research*, A299:544–547, 1990.

[62] I. G. Zubal, H. Tagare, L. Zhang, and J. Duncan. "3-D Registration of Intermodality Medical Images". In J. H. Nagel and W. M. Smith, editors, *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 293–294, 1991. vol. 13.