EDGE-PRESERVING MODELS AND EFFICIENT ALGORITHMS FOR ILL-POSED INVERSE PROBLEMS IN IMAGE PROCESSING

A Thesis
Submitted to the Faculty

of

Purdue University

by

Suhail S. Saquib

In Partial Fulfillment of the
Requirements for the Degree

of

Doctor of Philosophy

May 1997

To my parents and Afroz

ACKNOWLEDGMENTS

I would like to thank my advisor, Dr. Charles Bouman, for his boundless source of ideas, insight, and for inspiring me throughout the course of this work. He has been a valuable mentor, guide, and friend for the past few years and I have learned a lot from him. I can't thank him enough for "twisting my arm" to go to Los Alamos for my internship last summer.

This research work was done in collaboration with Dr. Ken Sauer, University of Notre Dame, and I would like to thank him for all his guidance, ideas, and suggestions. He was also responsible for obtaining all the tomographic data sets that were used in this research. I would also like to thank my committee members for their input and guidance.

It was a real positive experience working with Dr. Ken Hanson last summer in Los Alamos National Laboratories. Thanks are also due to Jeannene, Xavier, and Ann for making me feel at home in Los Alamos. I enjoyed having discussions with Greg Cunningham and Andreas Hielscher while I was in Los Alamos. I would like to thank Andreas Hielscher in particular for his assistance in setting up the simulation example in Chapter 5.

I would like to thank Trent Neel, Wright-Patterson Air Force Base, and Nicolas Dussausoy, Aracor, for providing the flashlight data. Thanks are also due to Tin-Su Pan and Michael A. King, University of Massachusetts, for providing the 3D SPECT data.

My stay in West Lafayette would have been significantly less enjoyable had it not been for my close group of friends. I acknowledge their support and their understanding. To Shahab, for being wild and funny, to Priya and Roxana for their great cooking, and to Manish for interesting discussions on the deeper aspects of life.

Recently, my life has become considerably richer and more meaningful due to Afroz. She has provided me with unfailing support ever since she entered my life. Last but not the least, as with all of my educational endeavors thus far, this thesis would not have been possible without the constant support of my parents. I dedicate this thesis to my parents and Afroz.

DISCARD THIS PAGE

TABLE OF CONTENTS

DISCARD THIS PAGE

LIST OF TABLES

DISCARD THIS PAGE

LIST OF FIGURES

Figure | Page

ABSTRACT

Saquib, Suhail S. Ph. D., Purdue University, May 1997. Edge-Preserving Models and Efficient Algorithms for Ill-Posed Inverse Problems in Image Processing.
Major Professor: Charles A. Bouman.

The goal of this research is to develop detail and edge-preserving image models to characterize natural images. Using these image models, we have developed efficient unsupervised algorithms for solving ill-posed inverse problems in image processing applications. The first part of this research deals with parameter estimation of fixed resolution Markov random field (MRF) models. This is an important problem since without a method to estimate the model parameters in an unsupervised fashion, one has to reconstruct the unknown image for several values of the model parameters and then visually choose between the results. We have shown that for a broad selection of MRF models and problem settings, it is possible to estimate the model parameters directly from the data using the EM algorithm. We have proposed a fast simulation technique and an extrapolation method to compute the estimates in a few iterations. Experimental results indicate that these fast algorithms substantially reduce computation and result in good parameter estimates for real tomographic data sets.

The second part of this research deals with formulating a functional substitution approach for efficient computation of the MAP estimate for emission and transmission tomography. The new method retains the fast convergence of a recently proposed Newton-Raphson method and is globally convergent.

The third part of this research deals with formulating non-homogeneous models. Non-homogeneous models have been largely ignored in the past since there was no

effective means of estimating the large number of model parameters. We have tackled this problem in the multiresolution framework, where the space-varying model parameters at any resolution are estimated from the coarser resolution image. Experimental results on real tomographic data sets and optical flow estimation results on real image sequences demonstrate that the multiresolution non-homogeneous model results in cleaner and sharper images as compared to the fixed resolution homogeneous model. Moreover, the superior quality is achieved at no additional computational cost.

The last part of this research deals with efficient image reconstruction from time-resolved diffusion data, which employs a finite-difference approach to solve the diffusion equation and adjoint differentiation to compute the gradient of the cost criterion. The intended application is medical optical tomography.

1. OVERVIEW

1.1 Introduction

Many image processing applications require the estimation of the underlying unknown image from noisy data that reflects some aggregate measurements of the object of interest. For instance, in computerized tomography (CT) applications the measurements consist of line integrals of the object of interest. In restoration applications involving blurring due to motion or incorrect focus, the measurements are low-pass filtered version of the original image.

Often the measurements that are collected are sparse and have a very low signal-to-noise ratio (SNR). This is a consequence of physical constraints that are encountered in practical applications. For example, in emission tomography where a radioactive substance is injected into the human body, one is limited by the amount of radiation exposure to the patient. The measurement time is also restricted when the internal organs of the patient are moving. Both of these constraints can degrade the quality of the measurements. In other instances the measurements are very sparse due to the constraints of the hardware system and/or the object structure. For example, in industrial CT applied to nondestructive inspection, the scanner cannot traverse completely around the object due to its size or due to the object being obstructed by other structure or because some rays are blocked entirely by X-ray opaque parts of the object. In other applications such as seismic exploration of the earth, the physical constraints on the placement of the sources and receivers restricts the number of the views that can be taken.

The problem of reconstructing the unknown image from sparse and noisy measurements is ill-posed and direct inversion methods that only employ a model for the measurements often yield inferior quality reconstructions. Bayesian methods on the other hand are more useful in this context since they incorporate a model for the underlying unknown image in addition to the data model. Consequently, the quality of the reconstructed image is substantially improved if the image model accurately describes the underlying unknown image. However, this approach could also degrade the quality of the reconstruction when there is a mismatch between the image model and the unknown image. This underlines the importance of formulating realistic image models that better describe natural images and incorporating parameters in the model that can be varied to reduce the model mismatch with the unknown image.

In practical applications, it is important to estimate these model parameters directly from the measurements to achieve the best possible results for each data set. In the absence of such an unsupervised scheme, one has to compute the reconstructions for several values of the model parameters and then visually choose between the results. This process is obviously very computationally expensive and therefore undesirable.

The first part of this thesis addresses the issue of unsupervised estimation of model parameters directly from the data. Fast algorithms are introduced that compute these estimates in a computationally efficient manner. The second part deals with efficient algorithms to compute tomographic reconstructions with desirable convergence properties. The third part deals with formulating a multiresolution non-homogeneous model that better characterizes natural images. The last part deals with an efficient model-based image reconstruction technique from time-resolved diffusion data, the intended application being medical optical tomography.

1.2 Outline

The following is an outline of each of the chapters in this thesis.

- Chapter 2:

  Markov random fields (MRF) have been widely used to model images in Bayesian frameworks for image reconstruction and restoration. Typically, these MRF models have parameters that allow the prior model to be adjusted for best performance. However, optimal estimation of these parameters (sometimes referred to as hyperparameters) is difficult in practice for two reasons: 1) Direct parameter estimation for MRF's is known to be mathematically and numerically challenging. 2) Parameters can not be directly estimated because the true image cross-section is unavailable.

In this chapter, we propose a computationally efficient scheme to address both these difficulties for a general class of MRF models, and we derive specific methods of parameter estimation for the MRF model known as generalized Gaussian MRF (GGMRF).

The first section of the chapter derives methods of direct estimation of scale and shape parameters for a general continuously valued MRF. For the GGMRF case, we show that the ML estimate of the scale parameter, σ, has a simple closed form solution, and we present an efficient scheme for computing the ML estimate of the shape parameter, p, by an off-line numerical computation of the dependence of the partition function on p.

The second section of the chapter presents a fast algorithm for computing ML parameter estimates when the true image is unavailable. To do this, we use the expectation maximization (EM) algorithm. We develop a fast simulation method to replace the E-step, and a method to improve parameter estimates when the simulations are terminated prior to convergence.

Experimental results indicate that our fast algorithms substantially reduce computation and result in good scale estimates for real tomographic data sets.

- Chapter 3:

  Statistical tomographic reconstruction algorithms generally require the efficient optimization of a functional. A recent algorithm known as iterative coordinate descent with Newton-Raphson updates (ICD/NR) has been shown to be much more computationally efficient than indirect optimization approaches based on the EM algorithm. However, while the ICD/NR algorithm has experimentally been shown to converge stably, no theoretical proof of convergence is known.

In this chapter, we prove that a modified algorithm, which we call ICD functional substitution (ICD/FS), has guaranteed global convergence in addition to the computational efficiency of ICD/NR. The ICD/FS method works by approximating the log-likelihood at each pixel by an alternative quadratic functional. Experimental results show that the convergence speed of the globally convergent algorithm is nearly identical to that of ICD/NR.

- Chapter 4:

  A good image model needs to be non-homogeneous to be able to adapt to the local characteristics of the different regions in an image. Toward this end, in this Chapter we propose a non-homogeneous Markov random field (MRF) model that has space-varying scale parameters. This formulation poses two difficulties: first, the number of scale parameters to estimate is on the order of the number of pixels in the image. Second, the scale parameters depend on the structure of the underlying image which is unknown. These two difficulties are solved by employing a generalized Gaussian MRF (GGMRF) based image model in a multiresolution framework. While the choice of the GGMRF enables us to estimate the local scale parameters in an intuitive fashion, the multiresolution framework yields two significant advantages: first, it makes it possible to estimate the space-varying scale parameters of the non-homogeneous MRF at any resolution by using the image at the coarser resolution. Second, it yields a multiresolution algorithm that is computationally efficient and more robust than its single resolution counterpart.

  Since the local scale parameters estimated from the coarser resolution image may over or under estimate the image variation by a fixed constant, we introduce a resolution dependent global scaling parameter in the model. These global scaling parameters are estimated directly from the data using the EM algorithm yielding a practical unsupervised reconstruction algorithm.

Experimental results on real tomographic data sets and optical flow estimation results on real image sequences demonstrate that the proposed non-homogeneous model and multiresolution reconstruction algorithm is superior to the homogeneous fixed resolution model in terms of quality of reconstruction and slightly better in terms of computational efficiency.

- Chapter 5:

  This chapter addresses the problem of reconstructing the unknown image of absorption and diffusion coefficients from time-resolved measurements of diffused light in a computationally efficient manner. The intended application is medical optical tomography, which has generated considerable interest in recent times. The inverse problem is posed in the Bayesian framework. The maximum *a posteriori* (MAP) estimate is used to compute the reconstruction. The diffusion model used for the measurements is solved forward in time using a finite-difference approach known as the alternating-directions implicit method. This method requires the inversion of a tridiagonal matrix at each time step and is therefore of $O(N)$ complexity, where N is the dimensionality of the image. Adjoint differentiation is used to compute the sensitivity of the measurements with respect to the unknown image. The novelty of our method lies in the computation of the sensitivity since we can achieve it in $O(N)$ time as opposed to $O(N^2)$ time required by the perturbation approach. Experimental results using simulated data show that the proposed method yields superior quality reconstructions with substantial savings in computation.

2. ML PARAMETER ESTIMATION FOR MARKOV RANDOM FIELDS

2.1 Introduction

Over the past decade, Bayesian methods for image reconstruction and restoration have become increasingly popular because they allow accurate modeling of both data collection, and image behavior. For example, Bayesian methods have been widely studied for emission tomography [1, 2, 3], transmission tomography [4, 5, 6], and image restoration [7, 8].

While the model used for data collection (also known as the forward model) has varied depending on the application, most of these approaches have used Markov random fields (MRF) to model the unknown image. This is because the MRF model is computationally tractable and can also capture many non-Gaussian aspects of images such as edges. A variety of continuously valued MRF models have been proposed for accurately modeling images [9, 2, 3, 10, 11, 12, 13]. Most of these are distinguished by the choice of potential function that assigns cost to differences between neighboring pixels.

While Bayesian methods can improve the quality of reconstructed images, they also have the potential to create errors when the models do not accurately characterize the data. To avoid this problem, modern approaches usually include parameters which allow the prior model and/or forward model to be adjusted to achieve the best possible results for each data set. Often the prior model parameters are referred to as hyperparameters because their effect is only indirectly apparent through the measured data.

Ideally, model parameters must be estimated for each data set as part of the image reconstruction or restoration process. However, estimation of these model parameters is often difficult for two reasons. First, direct maximum likelihood (ML) estimation of MRF parameters from example images is known to be a difficult problem. This is because in most cases the normalizing constant of the distribution, known as the partition function, is an intractable function of the parameters.

Second, in most applications example images are not available for estimation of parameters. Instead, parameters must be estimated indirectly from collected data because the true image cross-section is not known. This is a classic example of an incomplete data problem for which the expectation-maximization (EM) algorithm was developed [14, 15]. Intuitively, the EM algorithm works by iteratively computing the expectation of the unknown image statistics, and then maximizing the likelihood with respect to those statistics[1].

Most previous research for the direct estimation of MRF parameters has focused on discrete MRF's [17, 18, 19, 20, 21, 22] and used approximations to the ML estimate based on maximum pseudolikelihood [19, 20] or least squares [21]. The methods of these papers are not directly applicable to estimation of parameters from the continuously valued MRF's which are the subject of this Chapter. More recently, researchers such as Ogata, Geyer, Thompson, and Gidas have developed methods which attempt to directly compute the partition function during the estimation process [23, 24, 25, 26]. Since computation of the partition function requires on-line stochastic simulation, fast methods of simulation, such as those proposed by Potamianos, Goutsias, Jerrum and Sinclair, have been of considerable interest [27, 28].

Alternatively, estimation of parameters for Gaussian MRF's has received wide attention. This problem is essentially equivalent to estimation of regularization parameters in quadratic regularization. For a review of these approaches, see [29]. We do not consider these methods since, for the purposes of this research, Gaussian prior models lead to excessive smoothing of image edges.

[1] We note that this application of the EM algorithm is unrelated to Shepp and Vardi's classic use of EM in image reconstruction[16]. In Shepp and Vardi's work the image is treated as the unknown parameter.

A number of researchers have specifically studied the problem of estimating continuous MRF parameters from incomplete data. The simplest and perhaps most natural approach to this problem is joint MAP estimation of both the image and parameters [30]. Unfortunately, this leads to an inconsistent estimator that may even be divergent [31, 32]. Mohammad-Djafari has reported good results by using a joint MAP estimation method which is stabilized with an appropriate prior distribution for the parameters [33]. Schultz, Stevenson, and Lumsdaine have proposed a method for ML estimation of parameters by employing a signal-dependent approximation to the partition function and a quadratic approximation to the posterior distribution of the image [34]. Recently, Higdon *et. al.* have independently proposed a method for sampling from the posterior distribution of the MRF parameters using direct precomputation of the partition function [35]. Samples from the posterior distribution can be used to compute confidence intervals, and if the posterior distribution is peaked, then individual samples are likely to be good estimates.

In an effort to avoid the computationally expensive E-step of the EM algorithm, Zhou and Leahy have developed an approach which uses a mean field theory approximation to compute the required expectation [36, 32]. This method is philosophically similar to mean field approximations used by Zhang in segmentation problems [37]. Pun and Jeffs have taken an approach similar in concept to EM, but replacing the maximum likelihood step with a estimator designed specifically for the p parameter used in a Generalized Gaussian MRF (GGMRF) [38, 39].

True EM approaches have often been avoided due to the perception of excessive computation. One of the earliest attempts was by Geman and McClure when they proposed computing the shape parameter of a MRF by precomputing the partition function and replacing the expectation step with stochastic integration [1, 40].

In this Chapter, we propose a computationally efficient scheme to compute ML estimates of MRF model parameters from incomplete observations [41, 42, 43, 44].

Our method hinges on two innovations:

- Simple direct parameter estimation for continuous MRF's based on closed form expressions for the dependence of the partition function on the scale parameter.

- Computationally fast algorithms for computing the E-step of the EM algorithm based on fast simulation and parameter extrapolation.

In addition, we present a method for computing forward model parameters such as dosage which are often required for Bayesian reconstruction of transmission tomograms [45].

The first section of the Chapter derives methods for direct estimation of parameters for a general continuously valued MRF. To do this, we re-parameterize many well known potential functions using two parameters referred to as scale and shape. We show that in the general case, the scale parameter, σ, may be computed as the numerical solution to a simple equation. Interestingly, for the specific case of the generalized Gaussian MRF (GGMRF) [12], σ has a closed-form solution that is analogous to estimation of standard-deviation for Gaussian random variables [31]. Based on this result, we derive a general approach to ML estimation of the shape parameter through precomputation of an *one-dimensional function* of the shape parameter. We then illustrate the method for the case of the GGMRF by showing computed values of σ and p for a variety of images.

The second section of the Chapter presents a fast algorithm, based on the expectation maximization (EM) algorithm, for computing ML parameter estimates when the true image is unavailable. We develop a fast simulation method to replace the E-step based on extensions to the conventional Metropolis algorithm [46, 47, 48], and the heuristics suggested by Green and Han [49]. For the case of both transmission and emission tomography problems, our updates may be efficiently computed using the techniques described in [50]. To further reduce computation, we introduce a method to extrapolate the parameter estimates when the simulations are terminated prematurely.

Experimental results are presented for real transmission and emission data sets, as well as for image restoration. These results indicate that our fast algorithms substantially reduce computational cost and result in useful scale parameter estimates.

2.2 ML Parameter Estimation for Continuously Valued MRF's

Let X be a continuously valued unknown random image, and let Y be measured random observations. We use upper case letters to denote random quantities and lower case letters to denote their corresponding deterministic realizations. The unknown image is modeled by its probability density function, $\mathcal{P}_{\sigma,p}(x)$, where σ and p are unknown scale and shape parameters to be defined. We will assume that X contains N pixels indexed by X_s for $s \in S$, and that X takes values in the convex set $\Omega = \{x : x_i \geq 0 \text{ for all } i \in S\}$. The observations are modeled by their conditional density function, $\mathcal{P}(y|x)$. If Y is discrete, then $\mathcal{P}(y|x)$ is a probability mass function.

Our objective is to estimate the parameters σ and p so that we may compute the maximum *a posteriori* (MAP) estimate of X given Y

$$\hat{X}_{MAP} = \arg\max_{x\in\Omega} \{\log \mathcal{P}(Y|x) + \log \mathcal{P}_{\sigma,p}(x)\} \tag{2.1}$$

where the constraint of $x \in \Omega$ enforces positivity in the image. In this section, we develop methods for computing ML parameter estimates from direct observations of X.

2.2.1 MRF prior models

We model the unknown image X as an MRF with Gibbs distribution

$$\mathcal{P}_{\sigma,p}(x) = \begin{cases} \frac{1}{\sigma^N z(p)} \exp\left\{-\frac{1}{p} u(x/\sigma, p)\right\} & \text{if } x \in \Omega \\ 0 & \text{if } x \notin \Omega \end{cases} \tag{2.2}$$

where σ is a parameter that controls scale or variation in X, and p is a shape parameter that we will see appears in many common MRF models. We use the notation p because this is the traditional variable used in the GGMRF model which will be the focus of our analysis. The function $u(x/\sigma, p)$ is referred to as the energy function.

The normalizing constant of the distribution, $\sigma^N z(p)$, is known as the partition function, and is computed as $\int_{x\in\Omega} \exp\left\{-\frac{1}{p}u(x/\sigma,p)\right\} dx$. It is easily verified that this function is proportional to σ^N where N is the number of pixels.

We consider energy functions of the form

$$u(x/\sigma,p) = \sum_{\{i,j\}\in\mathcal{N}} b_{i-j}\rho\left(\frac{x_i - x_j}{\sigma}, p\right) \tag{2.3}$$

where $\mathcal{N}$ is the set of all neighboring pixel pairs, and $\rho(\cdot,\cdot)$ is the potential function which assigns a cost to differences between neighboring pixel values[2].

Depending on the choice of the potential function, (2.2) includes many common MRF models that have been proposed in the literature. Table 2.1 lists a variety of such potential functions. Notice that only the GGMRF model depends on p through the potential function. All other models listed depend on p solely through its explicit role in (2.2).

While these MRF models include Gaussian MRF's, Gaussian MRF's are of less interest because they tend to be excessively smooth. In an effort to better model image edges, a variety of functions have been suggested for $\rho(\Delta,p)$. These generally can be separated into functions which are convex or nonconvex in Δ. Convex potential functions often allow global optimization of (2.1) and are therefore computationally preferable. In addition, convex potential functions have been shown to lead to continuous or stable MAP estimates [12]. On the other hand, nonconvex potential functions tend to result in sharp discontinuities, which may be advantageous in applications such as edge detection.

The GGMRF model will be of particular interest to us because it will result in simple closed form expression for the ML estimate of σ. For this model, the density function for $X \in \Omega$ is given by

$$\mathcal{P}_{\sigma,p}(x) = \frac{1}{\sigma^N z(p)} \exp\left\{-\frac{1}{p\sigma^p} \sum_{\{i,j\}\in\mathcal{N}} b_{i-j}|x_i - x_j|^p\right\} \tag{2.4}$$

[2]Notice that this distribution is not proper since its integral is infinite. The distribution may be made proper by adding terms of the form $\epsilon\rho(x_i/\sigma,p)$ to the energy function. All results of the paper may then be obtained by taking the limit as $\epsilon \to 0$.

Table 2.1 List of nonconvex and convex potential functions that have been used.

Nonconvex potential functions		
Author (name)	Reference	$\rho(\Delta, p)$
Geman and McClure	[1, 40]	$\frac{\Delta}{1+\Delta^2}$
Blake and Zisserman (weak spring)	[9, 51]	$\min\{\Delta^2, 1\}$
Hebert and Leahy	[2]	$\log(1+\Delta^2)$
Geman and Reynolds	[11]	$\frac{\lvert\Delta\rvert}{1+\lvert\Delta\rvert}$
Convex potential functions		
Author (name)	Reference	$\rho(\Delta, p)$
(Gaussian)		Δ^2
Besag (Laplacian)	[52]	$\lvert\Delta\rvert$
Green	[3]	$\log\cosh\Delta$
Stevenson and Delp (Huber)	[10]	$\min\{\lvert\Delta\rvert^2, 2\lvert\Delta\rvert-1\}$
Bouman and Sauer (GGMRF)	[11]	$\lvert\Delta\rvert^p$

where normally $p \in [1, 2]$. Notice that (2.4) has a form which is analogous to a Gaussian distribution where σ plays the role of standard deviation. When $p = 2$, (2.4) reduces to a Gaussian model. Smaller values of p tend to produce sharper edges.

The GGMRF model has the advantage that its behavior is scale invariant [12, 13]. This property results from the fact that for all $x \in \Omega$, and $\sigma > 0$

$$u\left(x/\sigma, p\right) = \frac{1}{\sigma^p} u\left(x, p\right) \ . \tag{2.5}$$

While Bouman and Sauer [12] showed that the function $|x|^p$ characterized all scale invariant functions of $\Delta = x_i - x_j$, Brette, Idier, and Mohammad-Djafari [13] have shown that the class of scale invariant potential functions can be expanded if functions of both x_i and x_j are considered. For example, consider the divergence potential function proposed by O'Sullivan [53] and listed in Table 2.2. This function also leads to the scalable property of equation (2.5) for $p = 1$, and therefore results in a scale invariant prior. In addition, the divergence function is known to be a convex function of (x_i, x_j) [54]. A third example, which we call the generalized divergence, is given in Table 2.2. This function behaves like the divergence in the limit as $p \to 1$, but is well defined for $x_i = x_j = 0$. For a more detailed discussion of scale invariant priors, see [13].

2.2.2 ML estimation of σ

We first derive the ML estimate of the scale parameter, σ, in the general case, and then specialize it to the GGMRF or any other scale invariant MRF model which obeys the property of (2.5).

The normalized log-likelihood may be computed from (2.2) to be

$$\frac{1}{N} \log \mathcal{P}_{\sigma,p}(x) = \frac{-1}{Np} u(x/\sigma, p) - \log \sigma - \frac{1}{N} \log z(p) \ . \tag{2.6}$$

Table 2.2 List of scalable potential functions. The divergence and generalized divergence are examples of scalable potential functions which require two positive arguments.

Scalable potential functions			
Author (name)	Reference	$\rho(x_i, x_j, p)$	range of p
Bouman and Sauer (GGMRF)	[11]	$\|x_i - x_j\|^p$	$p > 0$
O'Sullivan (divergence)	[53, 13]	$(x_i - x_j)\log(x_i/x_j)$	p=1
(generalized divergence)	This Chapter	$(x_i - x_j)(x_i^{p-1} - x_j^{p-1})$	$p > 1$

Differentiating (2.6) with respect to σ and equating the result to zero yields the equation for $\hat{\sigma}$, the ML estimate of σ [41].

$$\frac{\hat{\sigma}}{Np}\frac{\partial}{\partial\sigma}u(x/\sigma,p)\bigg|_{\sigma=\hat{\sigma}} = -1 \tag{2.7}$$

While this expression may look complex, $\hat{\sigma}$ may be easily evaluated to any desired precision using a standard root finding algorithm such as half interval search. Perhaps this is somewhat surprising since in the general case ML estimation of MRF parameters is considered intractable due to the complex nature of the partition function. It is interesting to note that a similar parameterization by Ogata and Tanemura [55] did not lead to such a simple solution due to the assumption that $\mathcal{P}(x)$ had bounded support. In this case, the support of $\mathcal{P}(x)$ is dependent on σ and the result no longer holds.

For any scale invariant prior of Table 2.2, we may evaluate the expression of (2.7) by substituting in the scaling relation of (2.5). This results in the simple expression

$$\hat{\sigma}^p = \frac{1}{N}u(x,p)\ . \tag{2.8}$$

The above result is very appealing since it is quite simple, and applies for the GGMRF, divergence, and generalized divergence cases. In order to gain intuition, consider the case when X_i are i.i.d. Gaussian random variables. In this case $p=2$, σ^2 is simply the variance, and (2.8) reduces to the familiar expression $\hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{N} x_i^2$. Lange obtained a result equivalent to (2.8) in [31].

2.2.3 Joint ML estimate of σ and p for Scalable Priors

In this section, we will derive a method for computing the joint ML estimate of p and σ for the GGMRF model or any other model which obeys (2.5).

We can reduce this problem to a one-dimensional optimization since we have the closed form ML estimate of σ (2.8) in terms of p. To do this, we substitute the ML estimate of σ from (2.8) into the log-likelihood function of (2.6).

$$\frac{1}{N}\log\mathcal{P}_{\hat{\sigma},p}(x) = \frac{-1}{p} - \frac{\log(u(x,p)/N)}{p} - \frac{\log z(p)}{N} \tag{2.9}$$

The ML estimate of p is then given by

$$\hat{p} = \arg\min_p \left\{ \frac{\log(u(x,p)/N)}{p} + \frac{1}{p} + \frac{\log z(p)}{N} \right\} . \tag{2.10}$$

In this form, we can see that the function $\sigma^p = u(x,p)/N$ is a sufficient statistic for the parameter p.

The first two terms of (2.10) are easily computed, but the third term, $\frac{\log z(p)}{N}$, is generally an intractable function of p since it requires the computation of a N-dimensional integral. However, we next show that the derivative of this term may be computed using stochastic integration. The derivative is given by

$$\begin{aligned} \frac{d}{dp}\frac{\log z(p)}{N} &= \frac{d}{dp}\frac{1}{N}\log\int_{x\in\Omega}\exp\left\{-\frac{1}{p}u(x,p)\right\}dx \\ &= \frac{-1}{Nz(p)}\int_{x\in\Omega}\left(\frac{d}{dp}\frac{1}{p}u(x,p)\right)\exp\left\{-\frac{1}{p}u(x,p)\right\}dx \\ &= \frac{1}{p^2} - \frac{1}{Np}E\left[\left.\frac{d}{dp}u(X,p)\right|\sigma=1,p\right] \end{aligned} \tag{2.11}$$

where the last equality uses the consistency of the ML estimator for σ. The rest of the development in this section will be for the GGMRF prior. The extension to other scalable priors is similar.

Rewriting equation (2.11) for the GGMRF prior, we obtain

$$\frac{d}{dp}\frac{\log z(p)}{N} = \frac{1}{p^2} - \frac{1}{Np^2}\sum_{\{i,j\}\in\mathcal{N}} b_{i-j}\, E\left[|X_i - X_j|^p \log(|X_i - X_j|^p) | \, \sigma = 1, p\right]$$

where the function $\Delta^p log(\Delta^p)$ is interpreted to be 0 for $\Delta = 0$. Next, define the function $f(p)$ so that

$$\frac{df(p)}{dp} = \frac{-1}{Np^2}\sum_{\{i,j\}\in\mathcal{N}} b_{i-j}\, E\left[|X_i - X_j|^p \log(|X_i - X_j|^p) | \, \sigma = 1, p\right] . \tag{2.12}$$

Then the ML estimate of p is given by

$$\hat{p} = \arg\min_p \left\{ \frac{\log(u(x,p)/N)}{p} + f(p) \right\} . \tag{2.13}$$

The minimization of (2.13) may be evaluated by first computing $f'(p)$ of (2.12) using stochastic integration. The stochastic integration may be done by generating samples from the desired MRF using $\sigma = 1$, and computing the desired average. We note that this result rests on the reduction of the 2-D parameter estimation problem to a 1-D problem since in 1-D a derivative of $f'(p)$ is easily integrated to yield $f(p)$.

Note that all expectations are normalized by N. While the $\lim_{N\to\infty} \log \mathcal{P}_{\sigma,p}(x)$ generally does not exist, the normalized log-likelihood, $\lim_{N\to\infty} \frac{1}{N} \log \mathcal{P}_{\sigma,p}(x)$ does. Therefore, if we compute $f(p)$ for a sufficiently large lattice, we may assume that it does not vary with N.

2.2.4 ML estimate of σ and p for non-scalable priors

In this section, we derive methods to compute the joint ML estimates of σ and p when the potential function is not scalable. This includes all the potential functions of Table 2.1 except the Gaussian, Laplacian, and GGMRF.

Notice that $u(x,p)$ is not a function of p for any of the non-scalable potential functions. This means that for $\sigma = \hat{\sigma}$ the log-likelihood of (2.6) may be simplified to be

$$\frac{1}{N} \log \mathcal{P}_{\hat{\sigma},p}(x) = \frac{-1}{Np} u(x/\hat{\sigma}) - \log \hat{\sigma} - \frac{\log z(p)}{N} \ . \qquad (2.14)$$

where $\hat{\sigma}$ is given by (2.7). The term $\log z(p)/N$ may be computed in a manner similar to $f(p)$ by first computing its derivative.

$$\frac{d}{dp} \frac{\log z(p)}{N} = \frac{1}{Np^2} E\left[u(X) \mid \sigma = 1, p \right] \qquad (2.15)$$

Therefore the solution may be computed as the simultaneous solution to (2.7) and

$$\hat{p} = \arg\min_p \left\{ \frac{u(x/\hat{\sigma})}{Np} + \log \hat{\sigma} + \frac{\log z(p)}{N} \right\} \ .$$

2.3 Parameter Estimation from Incomplete Data

The previous section dealt with the problem of estimating the prior model parameters σ and p from the observed image X. However, in many applications the image X is never directly observed. For example, in tomography the photon counts, Y, are only indirectly related to the image X. In this case, there may be additional parameters, ϕ, related to the forward model, $\mathcal{P}_\phi(y|x)$.

Ideally, the ML parameter estimate is then given by

$$(\phi, \sigma, p) = \arg \max_{(\phi,\sigma,p)} \int_{x\in\Omega} \mathcal{P}_\phi(y|x)\mathcal{P}_{\sigma,p}(x)dx \quad . \tag{2.16}$$

While (2.16) is often difficult to compute directly, the EM algorithm is an effective method for iteratively maximizing (2.16) [14, 15].

In order to simplify notation, we will use the parameterization (γ, p) where $\gamma = \sigma^p$. Then a single update of the EM algorithm is given by

$$\phi_{k+1} = \arg \max_{\phi} E[\log \mathcal{P}_\phi(y|X)|Y = y, \phi_k, \gamma_k, p_k] \tag{2.17}$$

$$(\gamma_{k+1}, p_{k+1}) = \arg \max_{(\gamma,p)} E[\log \mathcal{P}_{\gamma,p}(X)|Y = y, \phi_k, \gamma_k, p_k] \tag{2.18}$$

where γ_k and p_k are the parameters generated at the k^{th} iteration of the EM algorithm. It can be shown that each iteration of the EM algorithm increases the likelihood, so that the likelihood value is guaranteed to converge to a local maximum.

For the GGMRF prior, the EM update of (2.18), may be explicitly computed as

$$p_{k+1} = \arg \min_{p} \left\{ \frac{1}{p} \log E\left[u(X, p)/N \mid Y = y,\ \phi_k, \gamma_k,\ p_k\right] + f(p) \right\} \tag{2.19}$$

$$\gamma_{k+1} = \frac{1}{N} E\left[u(X, p_{k+1}) \mid Y = y,\ \phi_k, \gamma_k,\ p_k\right] \tag{2.20}$$

The expectations of (2.19) and (2.20) may be approximated using **on-line** stochastic integration. This is done by generating samples from the posterior distribution of X given Y, and then computing the desired sample averages in place of the true expectations [1, 40, 22]. Unlike the off-line stochastic integration of (2.12), evaluation of EM updates must be computationally efficient. In Section 2.3.3, we will show how this is possible.

If p is known, then only σ needs to be estimated. In fact, estimation of σ is essential in many inverse problems. Too small a value of σ results in overly smooth images and too large a value of σ results in images with excessive noise. For this case, only (2.20) need be applied, and the expectation may be computed by averaging values of $u(X, p)$ for multiple samples of the image X generated from the posterior distribution of X given Y. We will discuss efficient algorithms for generating these samples in the next sections.

If both p and σ must be estimated, then the update of (2.19) must be computed first, and the result used to compute (2.20). Computation of (2.19) is somewhat more difficult since it requires that multiple samples of X be stored so the expectation may be computed as a function of p. However, we will show that often a single sample of X is sufficient to perform each EM update, so only a single image need be stored.

For non-scalable priors, the new parameters σ_{k+1} and p_{k+1} are given by the solution to the coupled equations

$$p_{k+1} = \arg\min_{p} \left\{ E\left[\frac{u(X/\sigma_{k+1})}{Np} \mid Y=y,\ \phi_k,\ \sigma_k,\ p_k \right] + \log \sigma_{k+1} + \frac{\log z(p)}{N} \right\} \quad (2.21)$$

$$\frac{\sigma_{k+1}}{Np_{k+1}} E\left[\frac{d}{d\sigma} u(X/\sigma)\Big|_{\sigma=\sigma_{k+1}} \mid Y=y,\ \phi_k,\ \sigma_k,\ p_k \right] = -1 \ . \quad (2.22)$$

These equations may be solved by iteratively computing the solution to each. Since each equation represents the minimization with respect to the corresponding variable, iterative solution will not diverge (if the ML estimate exists). When p is assumed known, the EM update for σ is given by (2.22) alone. However, computing the expectation of (2.22) requires buffering of the sample images.

2.3.1 Stochastic data models for tomography

In this section, we introduce the stochastic models that we will need for emission and transmission tomography. For a description of photon counting models in tomography see [16] and [4], and for a development which is notationally similar to the one presented here, see [50].

Let x denote the column vector of emission intensities in the emission case or the attenuation densities in the transmission case. For the emission case, let A_{ij} be the probability that a photon emitted from cell j is registered at the i^{th} detector. Let A be the projection matrix with elements $\{A_{ij}\}$, and let A_{i*} denote the i^{th} row of the projection matrix. Let y denote the column vector of measurements of Poisson-distributed photon counts at the detectors for all angles and displacements. Then, for the emission case, the log conditional distribution of the photon counts Y given x is

$$(emission) \quad \log \mathcal{P}(Y = y|x) = \sum_i \left(-A_{i*}x + y_i \log\{A_{i*}x\} - \log(y_i!)\right) . \tag{2.23}$$

This formulation is general enough to include a wide variety of photon-limited imaging problems, and the entries of A may also incorporate the effects of detector response and attenuation.

The transmission case is similar, but has A_{ij} corresponding to the length of intersection between the j^{th} cell and the i^{th} projection. Let the input photon counts be Poisson-distributed with rate y_T. Then the conditional log-likelihood of Y given x for the transmission case is

$$(transmission) \quad \log \mathcal{P}(Y = y|x) = \sum_i \left(-y_T e^{-A_{i*}x} + y_i(\log y_T - A_{i*}x) - \log(y_i!)\right) . \tag{2.24}$$

2.3.2 ML estimate of dosage y_T

The data for transmission tomography is often recorded in the form $z_i = \ln(\frac{y_T}{y_i})$. This preserves the ML estimates of integral densities, but results in the loss of the parameter y_T which is required for the log-likelihood of (2.24). If y_T is unknown, it can be estimated along with other parameters using the EM algorithm update equation of (2.17).

It is shown in Appendix A that the update equation for y_T is given as

$$y_{T\,k+1} = \frac{M}{2} \left\{ \sum_{i=1}^{M} E\left[\left\{ -e^{-A_{i*}X} - A_{i*}Xe^{-z_i} \right\} \mid Z = z,\ \sigma_k,\ p_k,\ y_{T\,k} \right] + e^{-z_i} + z_i e^{-z_i} \right\}^{-1} \tag{2.25}$$

2.3.3 Fast simulation technique

The EM parameter updates derived in Section 2.3 require the expectation of functions of X. Direct computation of these expectations is intractable, but we can approximate them by first generating sample images from the posterior distribution of X given Y and then computing averages using the sample images. The well-known

Metropolis algorithm [46] can be used to generate these samples from the posterior distribution, but it tends to suffer from slow convergence.

In this section, we propose a faster simulation method based on the algorithms of Hastings [47] and Peskun [48]. The experimental results indicate that the required expectations can be accurately estimated using only a single image sample.

Let $q(x'|x)$ be an arbitrary transition probability for generating a new state x' from the current state x. Then in order to generate a sample with distribution $\pi(x)$, one should accept new samples with probability

$$\alpha(x', x) = \min\left\{1, \frac{q(x|x')\pi(x')}{q(x'|x)\pi(x)}\right\} .$$

The Metropolis algorithm is a special case of this general formulation when we choose $q(x'|x) = q(x|x')$. Another special case is the Gibbs sampler [8] when the new state for pixel j is generated using the conditional distribution, under $\pi(x)$, of x_j given the values of all other pixels. For the Gibbs sampler, $\alpha(x, x') = 1$, and we always accept the new state.

A good choice of transition probability $q(x'|x)$ results in faster convergence of the stochastic simulation; but at present, the optimal selection of $q(x'|x)$ is an open problem. For the tomography problem, it has been shown that greedy pixel-wise optimization of the posterior distribution has fast convergence [6, 50]. We therefore conjecture that the Gibbs sampler is desirable for the tomography problem because each new pixel is generated from its marginal posterior distribution. Experimental results presented in Section 2.4.2 will support this conjecture.

Let us first examine the form of the conditional distribution of x_j required by the Gibbs sampler. Let x^n be the image at the n^{th} iteration. Then for the emission case, from (2.23), (2.2) and (2.3), we have

$$\begin{aligned}&\log \mathcal{P}(x_j|\{X_k = x_k^n : k \neq j\}, y) = \\ &\qquad \sum_i \left(-A_{ij}x_j + y_i \log\{A_{ij}(x_j - x_j^n) + A_{i*}x^n\}\right. - \frac{1}{p}\sum_{k\in\mathcal{N}_j} b_{j-k}\, \rho(\frac{x_j - x_k^n}{\sigma}, p) + C,\end{aligned} \tag{2.26}$$

where C is constant independent of x_j and $x_j > 0$. Note that directly generating samples from (2.26) would be very computationally expensive. Green and Han [49] suggested using a Gaussian distribution instead with parameters chosen to approximate the transition distribution of the Gibbs sampler. However, due to the non-Gaussian nature of our prior term, this approximation is good only for the data term [50] in (2.26). We can therefore obtain a good approximation by retaining the prior term as it is and using a second order Taylor series expansion for the data term of (2.26)

$$\log \mathcal{P}(x_j|\{X_k = x_k^n : k \neq j\}, y) \approx \\ d_1(x_j - x_j^n) + \frac{d_2}{2}(x_j - x_j^n)^2 - \frac{1}{p}\sum_{k\in\mathcal{N}_j} b_{j-k}\,\rho(\frac{x_j - x_k^n}{\sigma}, p) + C' \quad , \tag{2.27}$$

where d_1 and d_2 are the first and second derivative of the data term with respect to x_j evaluated at x_j^n. In [50] it is shown that

$$\begin{aligned} d_1 &= -\sum_i A_{ij}\left(1 - \frac{y_i}{\tilde{p}_i^n}\right) \\ d_2 &= -\sum_i y_i\left(\frac{A_{ij}}{\tilde{p}_i^n}\right)^2 \quad , \end{aligned}$$

where $\tilde{p}^n = Ax^n$. The approximation holds for the transmission case also, with the corresponding expressions for d_1 and d_2 as follows:

$$\begin{aligned} d_1 &= -\sum_i A_{ij}\left(y_i - y_T e^{-\tilde{p}_i^n}\right) \\ d_2 &= -\sum_i A_{ij}^2 y_T e^{-\tilde{p}_i^n} \quad . \end{aligned}$$

For efficient computation, we keep $\tilde{p}^n$ as a state vector and update it after each pixel update as follows

$$\tilde{p}^{n+1} = A_{*j}(x_j^{n+1} - x_j^n) + \tilde{p}^n \quad .$$

Let the transition distribution for generating the new state for the j^{th} pixel be denoted as $q_j(x)$. Then we would like $q_j(x)$ to be a Gaussian distribution with mode m

equal to the mode of the approximated conditional distribution (2.27). Unfortunately, generating positive samples from $q_j(x)$ is computationally intensive when $m \ll 0$. However, we can use the fact that the tail of a Gaussian distribution may be accurately approximated as an exponential distribution. Fig. 2.1 illustrates this fact. In the light of the above discussion, we choose $q_j(x)$ with the following form

$$q_j(x) = \begin{cases} \frac{1}{C(s,m)} \exp\left\{\frac{-(x-m)^2}{2s^2}\right\} & m > 0, x > 0 \\ \frac{1}{\beta} \exp\left\{\frac{-x}{\beta}\right\} & m < 0, x > 0 \\ 0 & x < 0 \end{cases} ,$$

where $C(s,m)$ is the normalizing constant of the truncated Gaussian distribution and m is the mode of the approximated conditional distribution (2.27)

$$m = \arg\max_x \left[d_1(x - x_j^n) + \frac{d_2}{2}(x - x_j^n)^2 - \frac{1}{p} \sum_{k\in\mathcal{N}_j} b_{j-k}\, \rho(\frac{x - x_k^n}{\sigma}, p) \right] . \qquad (2.28)$$

Choosing s^2 is more difficult due to the prior term. Since we can at best do an approximate fit to the original distribution (2.26), it is not clear whether a more precise choice of s^2 would yield a significant improvement in performance. We therefore choose

$$s^2 = \frac{1}{d_2}$$

to be the variance of the data term. Note that the variance of the approximated distribution (2.27) is over estimated by this particular choice. β is determined by setting the derivative of the log of the exponential distribution equal to the derivative of (2.27) at $x_j = 0$,

$$\beta = \left\{ -d_1 + d_2 x_j^n + \frac{1}{\sigma p} \sum_{k\in\mathcal{N}_j} b_{j-k}\, \rho'(\frac{-x_k^n}{\sigma}, p) \right\}^{-1} ,$$

where $\rho'(\cdot,\cdot)$ is the derivative of $\rho(\cdot,\cdot)$ with respect to its first argument.

Once d_1 and d_2 are computed, the optimization of (2.28) is computationally inexpensive since the sum associated with the prior typically involves few pixels. We use the half interval method to compute m. Note that during MAP reconstruction, m is also the updated value of pixel x_j in the iterative coordinate descent (ICD) algorithm of Bouman and Sauer [50].

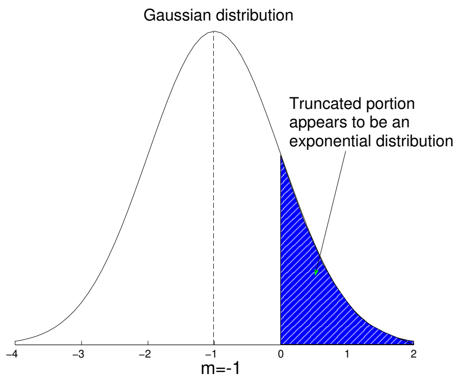


Fig. 2.1. The above figure shows that the truncated Gaussian distribution to ensure the positivity constraint can be approximated by an exponential distribution.

2.3.4 Extrapolation of parameter estimates

Even with exact computation of the E-step, the convergence of the EM algorithm can sometimes be slow. One way to further reduce the computation is to improve the current EM parameter estimates by extrapolating them. This extrapolation requires very little computation, so it may be applied at each EM iteration. At each iteration k, the extrapolated parameter, $\theta_k^{(e)}$, is then an improvement over the EM parameter θ_k.

First consider the case of the GGMRF where p is known and $\gamma = \sigma^p$ must be estimated. This is an important special case. It is well known that

$$\frac{d}{d\gamma} E\left[\log \mathcal{P}_\gamma(X) \mid Y = y, \hat{\gamma}\right]\Big|_{\gamma=\hat{\gamma}} = 0 \ ,$$

where $\hat{\gamma}$ is the maximum likelihood estimate of γ. From this it can be shown that

$$\hat{\gamma} = E\left[u(X,p)/N | Y = y, \hat{\gamma}\right] \ . \tag{2.29}$$

The EM algorithm iteratively solves for the fixed point of this equation. However, a fast method is to search directly for its root. Define the function

$$g(\gamma) = E\left[u(X,p)/N | Y = y, \gamma\right] - \gamma \ . \tag{2.30}$$

Then the ML estimate of γ is the solution to $g(\hat{\gamma}) = 0$. At iteration k of the EM algorithm, the value of $g(\gamma_k)$ is given by

$$\begin{aligned} g(\gamma_k) &= E\left[u(X,p)/N|Y=y,\gamma_k\right] - \gamma_k \\ &= \gamma_{k+1} - \gamma_k \ . \end{aligned} \tag{2.31}$$

Therefore, we can plot the function $g(\gamma)$ computed from (2.31).

Fig. 2.2 shows such a plot of $g(\gamma)$ with respect to γ for an emission phantom when we use only one sample of X to estimate the expectation of $u(X,p)$. At each iteration, we use the last three values of $(\gamma_k, g(\gamma_k))$ to compute a least squares fit to a straight line. The zero crossing of the least squares fit then yields the extrapolated value of $\gamma_k^{(e)}$. Note that $\gamma_k^{(e)}$ is close to the ML estimate after just 4 iterations.

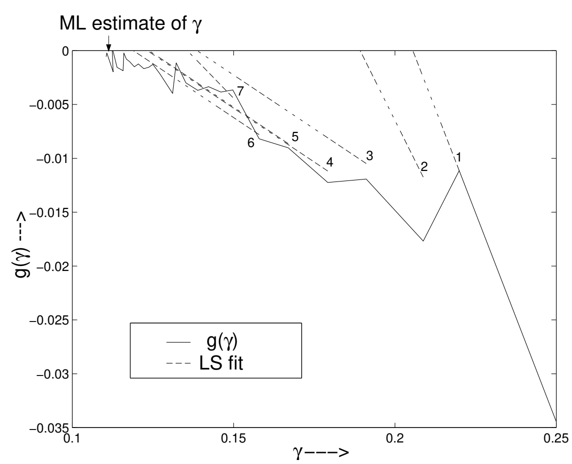


Fig. 2.2. The plot shows the normalized gradient $g(\gamma)$ computed at the EM updates of γ_k for a emission phantom using a GGMRF prior with $p = 1.1$. The ML estimate of γ is given by the root of $g(\gamma)$. The least squares (LS) fit obtained at the first 7 points are numbered (1-7) and shown with a dashed line. The intersection of the dashed lines with the top of the graph are the extrapolated parameter values, $\gamma_k^{(e)}$.

The generalization to the case when p is not known is conceptually easy. In this case, the ML estimates of γ and p are given as the roots of the following equations

$$\begin{aligned} \frac{d}{d\gamma} E[\log \mathcal{P}_{\gamma,\hat{p}}(X)|Y=y,\hat{\gamma},\hat{p}]\bigg|_{\gamma=\hat{\gamma}} &= 0 \\ \frac{d}{dp} E[\log \mathcal{P}_{\hat{\gamma},p}(X)|Y=y,\hat{\gamma},\hat{p}]\bigg|_{p=\hat{p}} &= 0 \ . \end{aligned}$$

Similar to the case where p is known, we can now define the vector valued function

$$g(\gamma,p)=\left[\begin{array}{l} E[u(X,p)/N|Y=y,\gamma,p]-\gamma \\ \frac{p}{N}E\left[\frac{d}{dp}u(X,p)-\frac{1}{p}u(X,p)\middle|Y=y,\gamma,p\right]+\gamma(1-\log\gamma)+\gamma p^2\frac{df(p)}{dp} \end{array}\right].$$

Then the ML estimates of γ and p are given as the roots of $g(\hat{\gamma},\hat{p})=0$. Note that we can easily compute $g(\gamma_k,p_k)$ when computing the EM updates for γ and p. The computed values of $g(\cdot,\cdot)$ at the past n EM updates are used to obtain least squares fits to two planes. The roots of the fitted planes are then the extrapolated values, $\gamma_k^{(e)}$ and $p_k^{(e)}$.

The non-scalable priors are handled in a similar fashion. In this case, the function $g(\cdot,\cdot)$ is given as

$$g(\sigma,p)=\left[\begin{array}{l} \frac{\sigma}{N}E\left[\frac{d}{d\sigma}u(X/\sigma)\mid Y=y,\ p,\ \sigma\right]+p \\ \frac{1}{N}E\left[u(X/\sigma)\mid Y=y,\ p,\ \sigma\right]-\frac{1}{N}E\left[u(X)\mid p,\ \sigma=1\right] \end{array}\right].$$

The two components of the vector valued function $g(\cdot,\cdot)$ can be rooted simultaneously or individually to obtain the extrapolated values, $\sigma_k^{(e)}$ and $p_k^{(e)}$.

2.4 Experimental Results

In the following two sections, we experimentally study the convergence speed and accuracy of the proposed parameter estimation method. Section 2.4.1 presents results of direct parameter estimation from observed images; while Section 2.4.2 presents results for parameter estimation from incomplete data.

2.4.1 Direct estimation of σ and p

In order to compute the ML estimates of p, we first computed the function $f'(p)$ using (2.12) and then integrated it using a second order spline to yield $f(p)$. To compute $f'(p)$, we computed batches of $10,000$ full iterations of a 64×64 periodic MRF with an 8 point neighborhood using $b_{i-j}=(2\sqrt{2}+4)^{-1}$ for nearest neighbors and $b_{i-j}=(4\sqrt{2}+4)^{-1}$ for diagonal neighbors. From each batch of $10,000$ iterations, an estimate of $f'(p)$ was computed. This procedure was repeated for a single value of

p until the estimate was found to stablize. Then the value of p was updated and the complete procedure repeated. Once the function $f(p)$ was computed, the accuracy of the result was tested by estimating p from sample GGMRF's with known values of p.

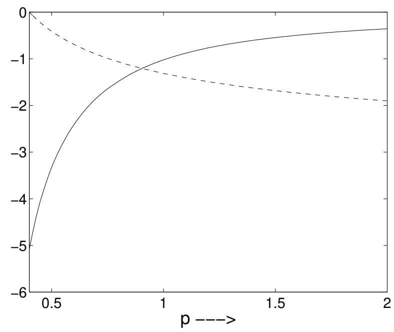


Fig. 2.3. The solid line shows $f'(p)$ and the dashed line shows $f(p)$.

Figure 2.3 shows the plots of $f(p)$ for $0.4 < p < 2.0$. Since the computation of $f(p)$ need only be done once, the speed of convergence is not a great issue. However, we found that more iterations were required as p decreased. Also, the function $f'(p)$ was sampled more finely for $p < 1$.

Figure 2.4 shows a host of natural and synthetic images with their corresponding joint ML estimates of p and σ. Figures 2.4(m) and (n) show two sample images generated using the GGMRF model with known values of p and σ. In both cases, the ML estimates are close to the true values.

Note that for most natural images except for a few texture images, the ML estimate of p was less than 1, and for many images containing man made objects p was less than 0.4. In fact, a similar result has been independently reported by Higdon *et. al.* [35]. Very small values of p may not lead to the best quality MAP reconstructions[3]. In particular, the tomographic cross-section in Fig. 2.4(o) yields value of $p < 0.4$ which we have found to be undesirable for MAP tomographic reconstruction. Since the ML estimator has well known optimality properties, this behavior of the ML estimate may be due to the mismatch between the typical tomographic cross-sections

[3]For p less than one, convergence of the MAP estimate can not generally be guaranteed since the functional being minimized is not convex.

and the GGMRF model. In light of this result, alternative methods for estimating p, such as those of Jeffs and Pun [39], might be advantageous depending on the intended application.

2.4.2 Estimation of σ and p from incomplete data

In this section, we study the performance of our proposed algorithms for estimating σ and p from incomplete observations. We present examples using both real and synthetic data for both tomographic reconstruction and image restoration. For the tomographic reconstruction examples, we fix $p = 1.1$ and estimate σ because the results of Section 2.4.1 indicate that ML estimates of p from tomographic cross-sections are excessively small. However, we emphasize that estimation of σ is a problem of primary importance in tomographic reconstruction since σ determines the overall smoothness of the reconstruction. MAP image reconstructions are then presented which indicate that the ML estimate of σ yields a good tradeoff between detail and noise reduction.

To illustrate the utility of our method for optimal joint estimation of p and σ, we apply our method to an image restoration problem, and show that for this case we can accurately estimate p and σ simultaneously from the noisy observations.

For tomographic reconstruction, we found that it was important to restrict the parameter estimates to the support of the object. If the flat background was included, then the estimation of σ tended to be too small and the reconstructions too smooth. For synthetic images, the support was known, but for real images it was extracted by first computing the convolution back projection (CBP) reconstruction, thresholding with a zero threshold, eroding 3 times, dilating 6 times and then eroding 3 times.

Figure 2.5 shows a synthetic emission phantom and the corresponding convolution back projection (CBP) reconstruction. The emission rates are on an array of 128 by 128 pixels of size 1.56mm^2, and 128 Poisson distributed projections are generated at each of 128 uniformly spaced angles. The total photon count was approximately 3 million.

Figure 2.6 shows the convergence of the ML estimate for σ using the GGMRF prior. We will refer to the three simulation methods as the conventional Metropolis (CM) method, the accelerated Metropolis (AM) method of Section 2.3.3, and the extrapolated-accelerated Metropolis (EAM) method of Section 2.3.4. For each case, each EM update is done after a single full sample of X is computed. The stochastic simulation is initialized with the CBP reconstruction and the EM algorithm is initialized with the ML estimate of σ obtained from the CBP reconstruction. Each plot also contains a line labeled as the true ML estimate. This value is computed by using 50 samples of X for each EM update and running the EM updates until convergence.

Figure 2.6(a) shows the results using the CM method with a transition distribution chosen to be Gaussian with the variance as the free parameter. Notice that the convergence rate varies substantially with the choice of variance. In practice, it is unclear how to choose the best variance before performing the simulations.

Figure 2.6(b) compares the EAM, AM and CM methods where the CM method uses the variance that produced the most rapid convergence. Notice that the EAM method has the most rapid convergence and all three methods converge to the desired ML value.

Figure 2.7 compares the quality of MAP reconstructions using $\hat{\sigma}$, $\hat{\sigma}/2$, $2\hat{\sigma}$, and $\hat{\sigma}_{CBP}$, the estimate obtained directly from the CBP. Of the four results, the ML estimate of σ seems to produce the most desirable tradeoff between detail and noise reduction. Figure 2.8 shows the corresponding reconstructions for the logcosh$(\cdot)$ prior with $p = 1$, 10, and 100 and the ML estimates of σ. The value $p = 10$ for the logcosh$(\cdot)$ prior yields reconstructions similar to that of a GGMRF prior with $p = 1.1$. The value of $p = 100$ for the logcosh$(\cdot)$ prior tends to the GGMRF with $p = 1$ and the MAP reconstruction is difficult to compute due to the extremely slow convergence.

Figure 2.9 shows the images corresponding to real transmission data for a flashlight. The original data consisted of 1024 projections taken at 792 equally spaced angles. From this original data, we generated a low resolution data set by retaining every 4^{th} projection at every 4^{th} angle. We used the full resolution data to form a "ground truth" image using CBP reconstruction which is shown in Fig. 2.9(a). All other reconstructions were then done with the lower resolution data. This approach allows us to determine if the reconstructions using the GGMRF prior actually produce more accurate detail.

Figure 2.9(b) shows the CBP reconstruction and Fig. 2.9(c) shows the GGMRF reconstruction using the ML estimate of the scale parameter. Figure 2.10 shows blowups of the same three images. Notice that the GGMRF reconstruction is sharper than the CBP reconstruction, and in some regions, it contains more detail than the ground truth image reconstructed with 16 times more data.

Figure 2.11(a) compares the EM updates of σ for the flashlight data using the CM, AM and EAM methods. It seems from this plot that the estimate obtained from the CM method has a bias. But Fig. 2.11(b) shows that the after a large number of iterations, the CM method tends toward the ML estimate. Figure 2.12(a) shows the support of the flashlight that was used to compute the ML estimate.

Figure 2.13 shows the reconstructions corresponding to a 3D SPECT data set obtained from cardiac perfusion imaging using Tc-99m sestamibi. For each slice, 128 projections were taken at 120 uniformly spaced angles between 0 and 2π. Figure 2.13(a) shows the CBP reconstruction of one slice. The reconstruction was done at 128 by 128 0.356cm pixel resolution. The total photon count for this slice was 148761. Figures 2.13(b-d) compare the MAP reconstructions corresponding to different values of the scale parameter σ. Again we see that the ML estimate of σ produces a reasonable tradeoff between detail and noise reduction.

Figure 2.14 shows the parameter estimation plots using the CM, AM and EAM method for the SPECT data. In this case it takes just 1 iteration for the AM or EAM method to converge to the ML estimate whereas the CM method takes about 15 iterations. Figure 2.12(b) shows the support of the object that was used to compute the ML estimate.

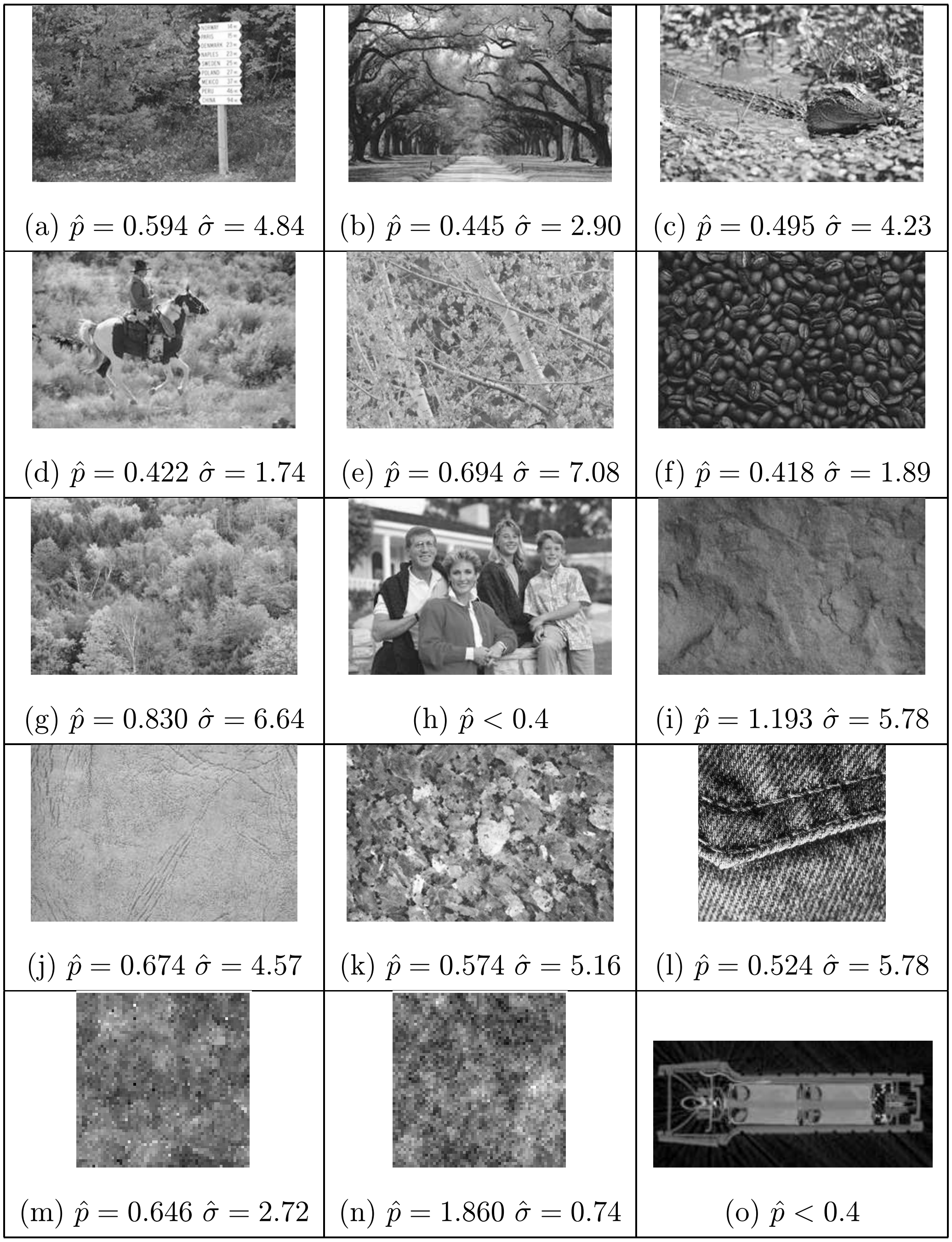
Figure 2.15(a) shows the original texture image that we use for a restoration example. Figure 2.15(b) shows the noisy image obtained by adding uncorrelated

Gaussian noise. The SNR of the observations is 37dB. Fig. 2.16 shows the joint estimation of p and σ for this case using the CM, AM and EAM methods. Note that the ML estimates obtained are very close to the estimates obtained from the original image. It takes about 10 iterations for the AM and EAM methods to converge to the ML estimate as compared to 20 iterations for the CM method. Figure 2.15(c) shows the MAP restoration for this example using the ML estimates of p and σ.

2.5 Conclusion

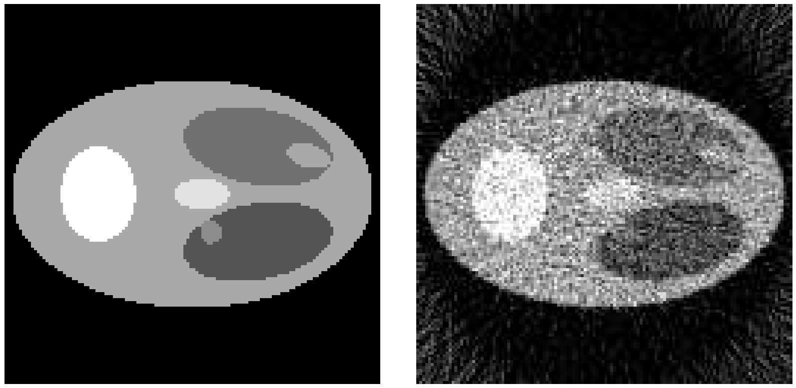
We have shown in this Chapter that maximum-likelihood estimation of free parameters for Bayesian image reconstruction is feasible for a broad selection of image models and problem settings. Our method is based on parameterization of continuous MRF's by a scale parameter, σ, and a shape parameter, p. For the class of scalable MRF's, the ML estimate of σ may be easily computed in closed form. For other continuous MRF's, the ML estimate of σ may be easily computed as the solution to an equation. Using this result, we also derive a method for computing the ML estimate of the shape parameter, p.

In most practical problems, σ and p must be estimated indirectly from measured data. For this case, we employ the EM algorithm, and develop a fast simulation algorithm together with a method for extrapolating the estimates when the EM algorithm is prematurely terminated. Together these methods allowed good parameter estimates to be computed in less than 10 iterations for the real and synthetic data sets that were used.



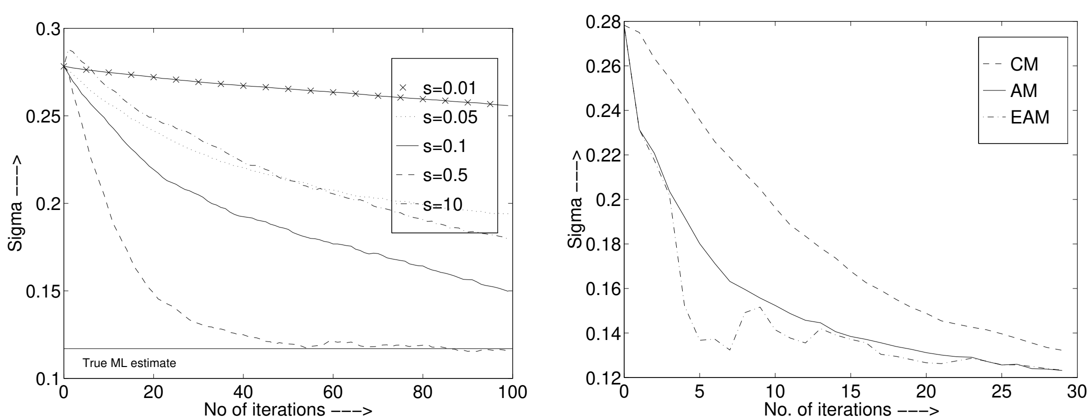
(a) $\hat{p} = 0.594$ $\hat{\sigma} = 4.84$
(b) $\hat{p} = 0.445$ $\hat{\sigma} = 2.90$
(c) $\hat{p} = 0.495$ $\hat{\sigma} = 4.23$
(d) $\hat{p} = 0.422$ $\hat{\sigma} = 1.74$
(e) $\hat{p} = 0.694$ $\hat{\sigma} = 7.08$
(f) $\hat{p} = 0.418$ $\hat{\sigma} = 1.89$
(g) $\hat{p} = 0.830$ $\hat{\sigma} = 6.64$
(h) $\hat{p} < 0.4$
(i) $\hat{p} = 1.193$ $\hat{\sigma} = 5.78$
(j) $\hat{p} = 0.674$ $\hat{\sigma} = 4.57$
(k) $\hat{p} = 0.574$ $\hat{\sigma} = 5.16$
(l) $\hat{p} = 0.524$ $\hat{\sigma} = 5.78$
(m) $\hat{p} = 0.646$ $\hat{\sigma} = 2.72$
(n) $\hat{p} = 1.860$ $\hat{\sigma} = 0.74$
(o) $\hat{p} < 0.4$

Fig. 2.4. The above figure shows joint ML estimation of p and σ for a variety of different natural images. In particular, (a-h) are images of natural scenes, (i-l) are images of different textures, (m) and (n) are synthetic images generated from the GGMRF distribution with parameters $p = 0.6,\ \sigma = 2.34$ and $p = 1.8,\ \sigma = 0.72$ respectively, and (o) is the CBP image obtained from transmission data for a flashlight.



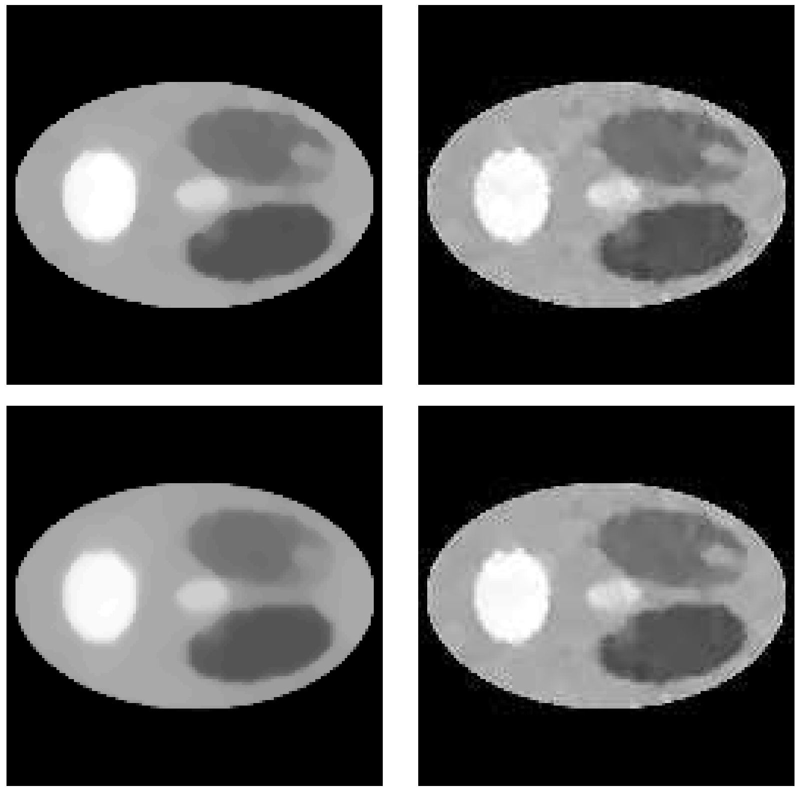
(a) (b)

Fig. 2.5. (a) Original emission phantom and (b) convolution back projection (CBP) reconstruction.



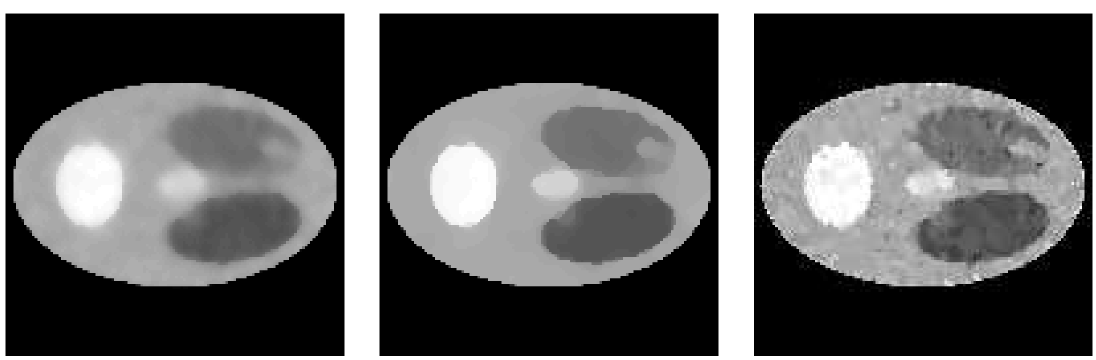
(a) (b)

Fig. 2.6. Convergence plots of σ for the emission phantom modeled by a GGMRF prior ($p = 1.1$). (a) CM method where s denotes the standard deviation of the symmetric transition distribution. (b) EAM method, AM method, and the CM method. All the updates are done using a single sample of X to compute the expectation. However, the true ML estimate is the converged value of σ when 50 samples are used to compute the expectation.



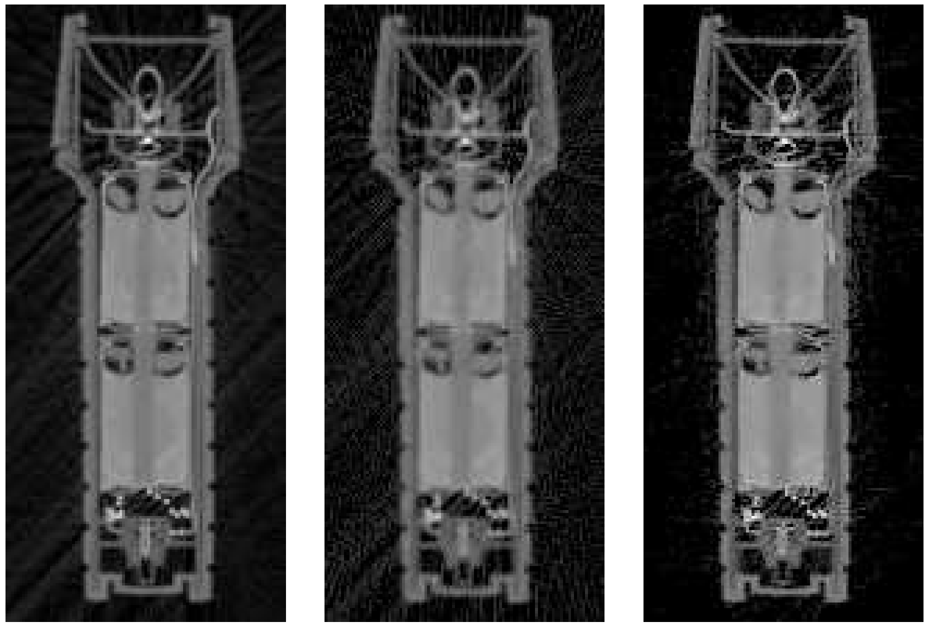
(a) (b) (c) (d)

Fig. 2.7. Reconstructed emission phantom using GGMRF prior with $p = 1.1$. The scale parameter σ is (a) $\hat{\sigma}$, (b) $\hat{\sigma}_{CBP}$, (c) $\frac{1}{2}\hat{\sigma}$, and (d) $2\hat{\sigma}$.



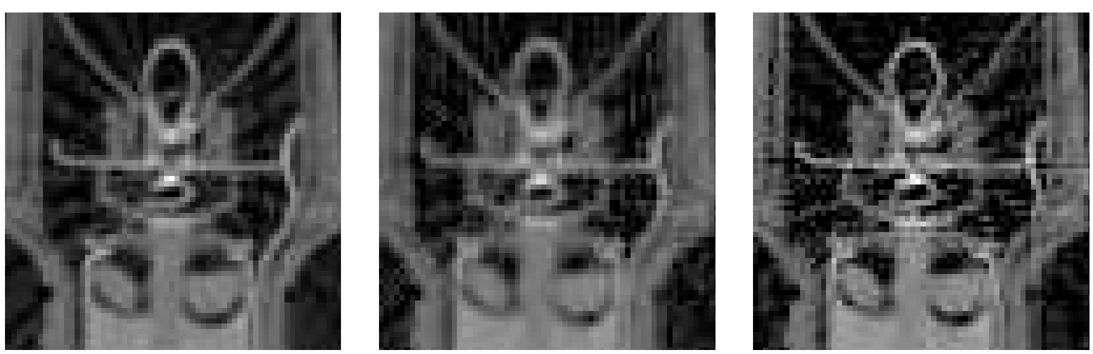
(a) (b) (c)

Fig. 2.8. Reconstructed emission phantom using logcosh$(\cdot)$ prior with the scale parameter σ optimally estimated for different values of p. The value of p is (a) 1, (b) 10, and (c) 100.



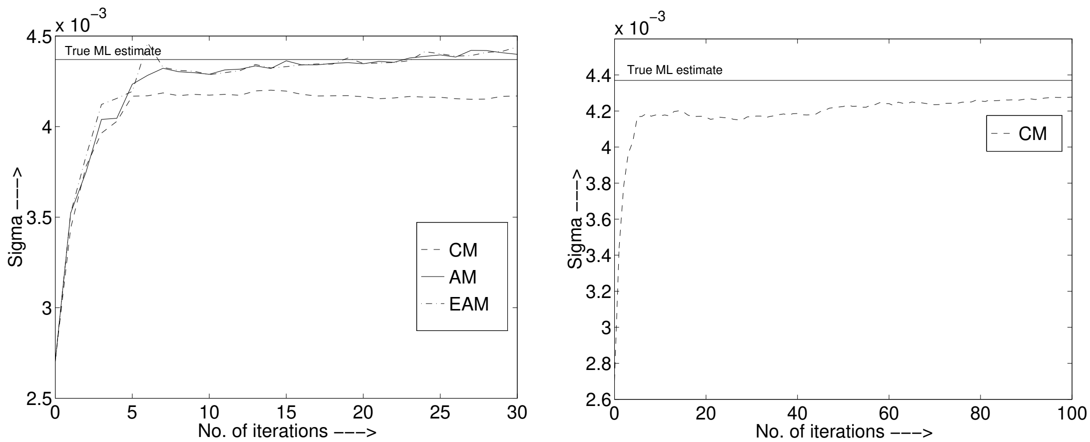
(a) (b) (c)

Fig. 2.9. (a) Ground truth obtained from high resolution transmission data, (b) CBP image, and (c) Reconstructed image using GGMRF prior with $p = 1.1$ and $\sigma = \hat{\sigma}$. (Data courtesy of Trent Neel, Wright-Patterson Air Force Base, and Nicolas Dussausoy, Aracor.)



(a) (b) (c)

Fig. 2.10. Blowup of images corresponding to Fig. 2.9. (a) Ground truth, (b) CBP image, and (c) Reconstructed image using GGMRF prior with $p = 1.1$ and $\sigma = \hat{\sigma}$.



(a) (b)

Fig. 2.11. Comparison of σ's convergence for CM, AM, EAM algorithms for the flashlight data modeled by a GGMRF prior ($p = 1.1$). The true ML estimate is the converged value of σ when 50 samples are used to compute the expectation.

(a) (b)

Fig. 2.12. These binary images show the support used for parameter estimation for the a) flashlight data and b) 3-D SPECT data set.

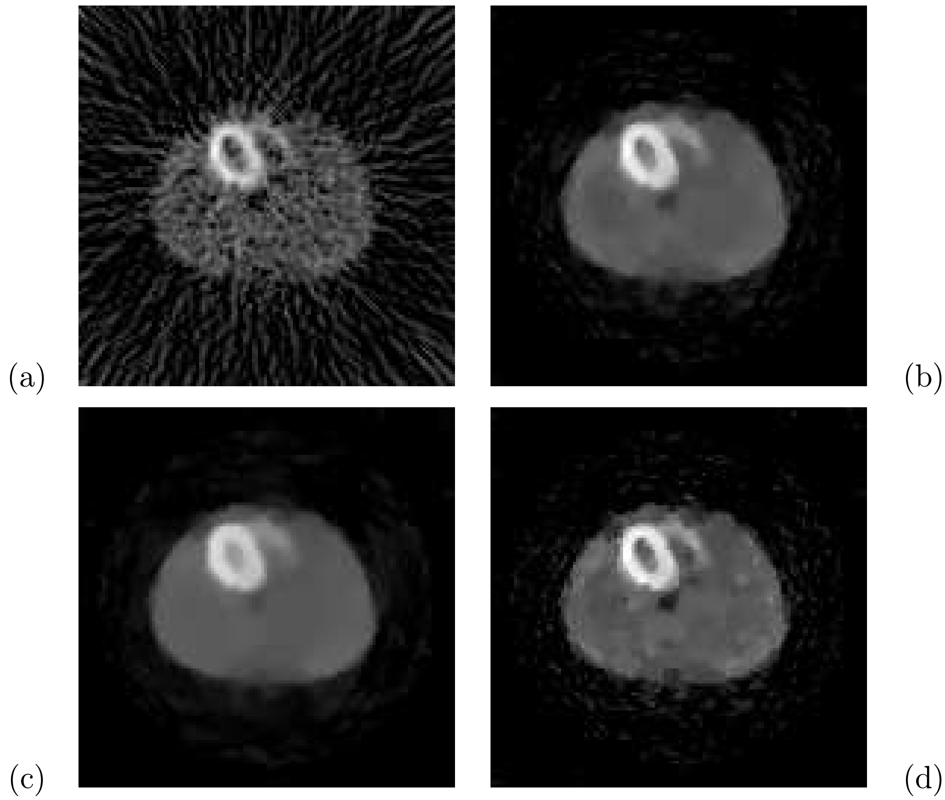


Fig. 2.13. (a) CBP reconstruction; Reconstructions using GGMRF prior with $p = 1.1$ and (b) $\sigma = \hat{\sigma}$, (c) $\sigma = \hat{\sigma}/2$, (c) $\sigma = 2\hat{\sigma}$. (Data courtesy of Tin-Su Pan & Michael A. King, Univ. of Massachusetts.)

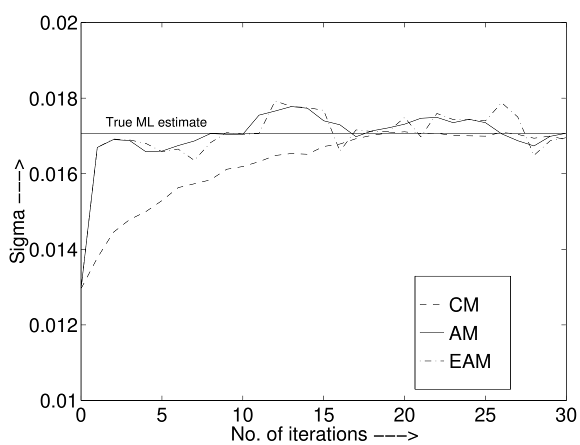
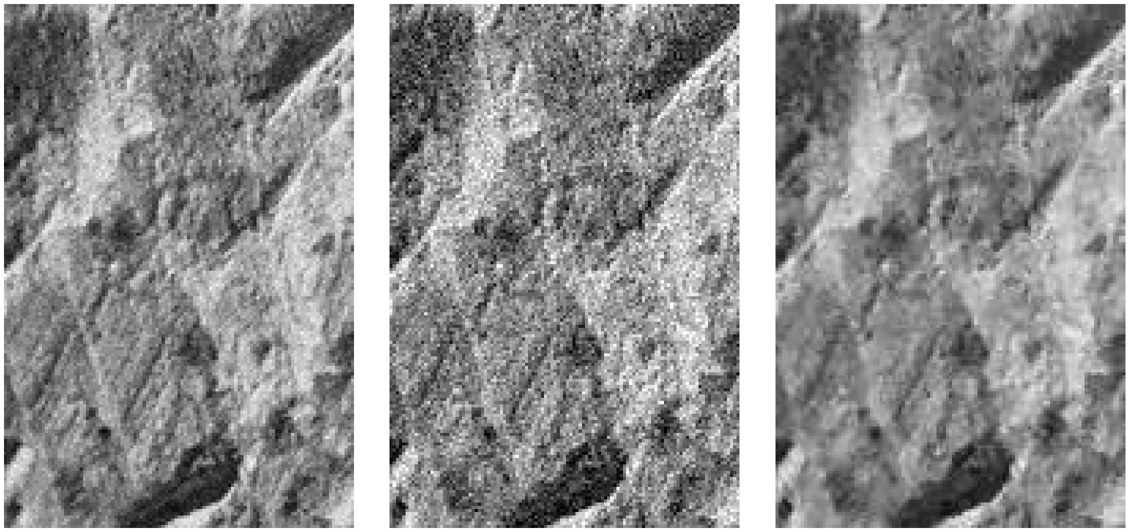
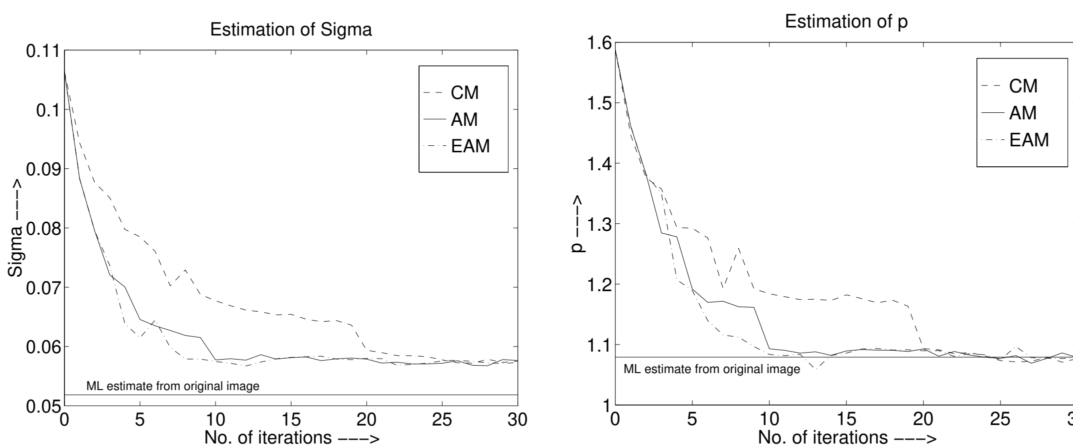


Fig. 2.14. Comparison of σ's convergence for CM, AM, EAM algorithms for the SPECT data modeled by a GGMRF prior ($p = 1.1$). The true ML estimate is the converged value of σ when 50 samples are used to compute the expectation.



(a) (b) (c)

Fig. 2.15. (a) Original image, (b) image corrupted with Gaussian noise (37dB), and (c) restored image using GGMRF prior and ML estimates of σ and p.



(a) (b)

Fig. 2.16. These plots show the EM updates for (a) σ, and (b) p for the restoration example using a GGMRF prior. The plots also show the ML estimate obtained for σ and p using the original image. All the updates are done using a single sample of X to compute the expectation.

3. PROVABLY CONVERGENT COORDINATE DESCENT IN STATISTICAL TOMOGRAPHIC RECONSTRUCTION

3.1 Introduction

Statistical tomographic reconstruction methods such as maximum-likelihood (ML) and maximum *a posteriori* probability (MAP) estimation seek the solution that best matches the probabilistic behavior of the data. But these estimates may require excessive computation for the resulting large-scale, iterative optimizations. The expectation-maximization (EM) algorithm has been widely applied for computing the ML estimates in the emission tomography problem [16]. Several other algorithms take an indirect approach based on EM to solve the MAP optimization problem[2, 56]. But these algorithms retain one main problem of EM, i.e. its slow convergence speed for tomographic reconstructions.

Recently, Bouman and Sauer have proposed an iterative coordinate descent (ICD) algorithm [50] for statistical image reconstruction based on the direct optimization of the MAP criterion. This algorithm is based on the sequential greedy optimization of pixel values in the reconstruction. It applies a truncated Taylor series expansion to derive a local quadratic approximation to the exact log-likelihood function. We use the term iterative coordinate descent/Newton Raphson (ICD/NR) for this computationally advantageous version of the coordinate descent algorithm. It has been experimentally demonstrated to converge very rapidly compared to EM algorithms, but is thus far not guaranteed theoretically to converge to the unique global MAP solution.

In this Chapter, we present a modified ICD algorithm which we call ICD functional substitution (ICD/FS). The ICD/FS algorithm locally approximates the exact log-likelihood function with an alternative quadratic function. We prove that the ICD/FS algorithm is globally convergent for both the emission and transmission reconstruction problem when the log-prior distribution is convex. As with ICD/NR, the new method easily incorporates non-negativity constraints and non-Gaussian prior distributions. We also note that as a functional substitution approach, the new algorithm is mathematically similar to optimization methods allowing parallel updates of arbitrary sets of pixels within the sequential framework [57, 58].

We include experimental results on synthetic phantoms, using both ICD/NR and ICD/FS algorithms for the comparison under the emission and transmission reconstruction cases. These results indicate that ICD/FS retains the rapid convergence properties of ICD/NR, but with the desirable global convergence properties.

3.2 Modeling Assumptions

The MAP tomographic reconstruction problem results in the numerical optimization problem of

$$x_{MAP} = \arg\max_x \{\log P(Y=y|X=x) + \log P(X=dx)\} \quad ,$$

where x is the unknown image and y is the projection data. We refer to $\log P(Y = y|X = x)$ as the data term, and the density function $\log P(X = dx)$ as the prior term. The approach we propose is based on the ICD algorithm. In this Chapter, we will assume that the function $\log P(X = x)$ is strictly concave with continuous derivatives on the set $\mathbb{R}^{N+} = \{x \in \mathbb{R}^N : x_j \geq 0\}$. For the emission case, we have

$$\log P(Y=y|X=x) = \sum_{i=1}^{M} \left(-A_{i*}x - r_i + y_i \log\{A_{i*}x + r_i\} - \log(y_i!)\right) \quad , \tag{3.1}$$

where M is the number of projections, A_{i*} is the i^{th} row of the projection matrix A, y_i are the observed photon counts for projection i, and r_i are additive terms usually due to background noise or random coincidences in the case of PET. For the transmission case, we have

$$\log P(Y=y|X=x) = \sum_{i=1}^{M} \left(-y_T \exp^{-A_{i*}x} + y_i(\log y_T - A_{i*}x) - \log(y_i!) \right) \quad , \qquad (3.2)$$

where y_T is the photon dosage per ray [50].

We will prove global convergence by verifying that ICD/FS meets the conditions of the theorem presented by Fessler and Hero in [59] for convergence of space-alternating generalized EM. Since this proof requires continuity of the log-likelihood on $\mathbb{R}^{N+}$, we must assume that the background noise is greater than zero, i.e. $r_i > 0$. We discuss alternative methods for the case $r_i = 0$ later in the Chapter.

3.3 The Functional Substitution Approach to ICD

In this section, we will develop the ICD/FS algorithm.

3.3.1 ICD/FS algorithm

The ICD method sequentially optimizes with respect to each pixel (i.e. coordinate of x). Let x^n be the image at the n^{th} iteration. Then the ICD update of the pixel x_j is computed by solving the equation

$$x_j^{n+1} = \arg\min_{x_j \geq 0} \{F_j(x_j) + P_j(x_j)\} \quad , \qquad (3.3)$$

where $F_j(x_j)$ and $P_j(x_j)$ represent the contribution of the data and prior terms, respectively, to the objective function expressed in terms of x_j. The particular form of $F_j(x_j)$ varies for the emission and transmission tomography cases of equations (3.1) and (3.2). But in both cases $F_j(x_j)$ is a convex function on $\mathbb{R}^{N+}$ and its derivative $f_j(x_j) = \frac{dF_j(x_j)}{dx_j}$ is strictly concave.

Unfortunately, direct optimization of (3.3) is not desirable because evaluation of the data term, $F_j(x_j)$, is computationally expensive. ICD/NR solves this problem by approximating the function with its second order Taylor series, replacing $F_j(x_j)$ with

$$F_j^{(nr)}(x_j) = \theta_1^{(nr)}(x_j - x_j^n) + \frac{1}{2}\theta_2^{(nr)}(x_j - x_j^n)^2 \quad ,$$

where $\theta_1^{(nr)} = f_j(x_j^n)$, $\theta_2^{(nr)} = \frac{df_j(x_j^n)}{dx_j}$. Although this approximation has been shown experimentally to be quite good, a theoretical proof of convergence for the resulting iterations has not been found.

The new algorithm ICD/FS results from using a slightly different value for θ_2. This new choice of θ_2 is slightly more conservative and will allow us to prove the desired global convergence property. The update equations for ICD/FS are given as follows:

$$\theta_1^{(fs)} = f_j(x_j^n) \tag{3.4}$$

$$\theta_2^{(fs)} = \begin{cases} \frac{f_j(x_j^n)-f_j(0)}{x_j^n} & \text{if } x_j^n > 0 \\ \frac{df_j(0)}{dx_j} & \text{if } x_j^n = 0 \end{cases} \tag{3.5}$$

$$F_j^{(fs)}(x_j) = \theta_1^{(fs)}(x_j - x_j^n) + \frac{1}{2}\theta_2^{(fs)}(x_j - x_j^n)^2 \tag{3.6}$$

$$x_j^{n+1} = \arg\min_{x_j \geq 0} \left\{ F_j^{(fs)}(x_j) + P_j(x_j) \right\} \tag{3.7}$$

The ICD/FS algorithm can be applied in both emission and transmission tomography problems. The only difference between these two cases is the specific computation of the values for $f_j(x_j^n)$, $f_j(0)$, and $\frac{df_j(0)}{dx_j}$ in (3.4) and (3.5). For the emission case, these values are given by

$$f_j(x_j^n) = \sum_{i=1}^{M} A_{ij}\left(1 - \frac{y_i}{p_i^n}\right) \tag{3.8}$$

$$f_j(0) = \sum_{i=1}^{M} A_{ij}\left(1 - \frac{y_i}{p_i^n - A_{ij}x_j^n}\right) \tag{3.9}$$

$$\frac{df_j(0)}{dx_j^n} = \sum_{i=1}^{M} y_i \left(\frac{A_{ij}}{p_i^n - A_{ij}x_j^n}\right)^2 , \tag{3.10}$$

where A_{ij} is the contribution of the j^{th} pixel to the i^{th} projection, and $p_i^n = A_{i*}x^n + r_i$ is the i^{th} projection of the reconstruction at iteration n. Note that p_i^n may be efficiently updated by $p_i^{n+1} = p_i^n + A_{ij}(x_j^{n+1} - x_j^n)$, with computation reduced by the sparse structure of A.

For the transmission case, the update values are given by

$$f_j(x_j^n) = \sum_{i=1}^{M} A_{ij}\left(y_i - y_T e^{-p_i^n}\right) \tag{3.11}$$

$$
\begin{aligned}
f_j(0) &= \sum_{i=1}^{M} A_{ij}\left(y_i - y_T e^{-p_i^n} e^{A_{ij}x_j^n}\right) && (3.12)\\
\frac{df_j(0)}{dx_j^n} &= \sum_{i=1}^{M} A_{ij}^2 y_T e^{-p_i^n} e^{A_{ij}x_j^n} \quad , && (3.13)
\end{aligned}
$$

where $p_i^n = A_{i*}x^n$.

ICD/FS has essentially the same computational requirements as ICD/NR since it generally requires the computation of two first derivatives in place of the first and second derivatives required for ICD/NR.

3.3.2 Global convergence of ICD/FS

In order to prove the global convergence of this new algorithm, we simply verify that it meets the two assumptions and six conditions of the global convergence proof presented in [59].

Most of these conditions are either the same as for [59], or are may be simply verified [1]. However, we will demonstrate the critical Condition 1, which states that the change in the substitute function is an upper bound on the change in the true functional to be minimized.

By the construction of function $f_j^{(fs)}(x)$, we know that $f_j^{(fs)}(0) = f_j(0)$, and $f_j^{(fs)}(x_j^n) = f_j(x_j^n)$. Since for both the emission and transmission case, $f_j(x)$ is a concave function and $f_j^{(fs)}(x)$ is a linear function, it follows that

$$
f_j(x) \begin{cases} \geq f_j^{(fs)}(x) & 0 \leq x < x_j^n \\ \leq f_j^{(fs)}(x) & x > x_j^n \end{cases} .
$$

Integration of $f_j(x)$ and $f_j^{(fs)}(x)$ result in the inequality

$$
F_j(x) - F_j(x_j^n) \leq F_j^{(fs)}(x) - F_j^{(fs)}(x_j^n) \; .
$$

Defining the functions $\Phi_j(x) = F_j(x) + P_j(x)$ and $\Phi_j^{(fs)}(x) = F_j^{(fs)}(x) + P_j(x)$ then results in the following lemma.

[1] Continuity of $f_j^{(fs)}(t)$ as a function of (t, x) on $\mathbb{R}^{(N+1)+}$ also appears to be necessary in Condition 2.

Lemma: Let $F_j(x)+P_j(x)$ be convex, and $F_j(x)$ be continuously differentiable on $I\!R^{N+}$. Furthermore, let $f_j(x)=\frac{dF_j(x)}{dx}$ be concave and continuous on $I\!R^{N+}$, and let

$$x_j^{n+1} = \arg\min_{x\geq 0}\{F_j^{(fs)}(x)+P(x)\}.$$

Then for all $x \in I\!R^{N+}$

$$\Phi_j(x)-\Phi_j(x_j^n) \leq \Phi_j^{(fs)}(x)-\Phi_j^{(fs)}(x_j^n)\ .$$

Based on this lemma and the conditions proved in [59], the global convergence of the ICD/FS algorithm follows.

3.3.3 ICD/FS with zero background emission noise

As mentioned previously, the emission case when $r_i=0$ is special since in this case the log-likelihood term may tend to $-\infty$ on $I\!R^{N+}$. This occurs in the unusual case in which x_j is the only nonzero pixel on a projection which has a nonzero photon count. In this case, $A_{i*}x = A_{ij}x_j$ and the log-likelihood functions have terms of the form $\log x_j$ which tend to $-\infty$ as $x_j \to 0$. There are a number of possible strategies for handling this case numerically.

Strategy 1: A very simple method for handling this case is to set r_i to a very small number such as $r_i=\frac{1}{100M}$. This guarantees that the expected number of additional photons due to this adjustment summed over all projections is much less than 1. In practice, such a small perturbation to the model should not have a significant effect on the resulting reconstruction. This strategy also has the added benefit of making the algorithm more robust to floating point round-off error.

Strategy 2: Modify the algorithm so that in the case when $f_j(0)=-\infty$, the function is recomputed at $f_j(\epsilon)$ where ϵ is chosen to assure that $f_j(\epsilon)<0$. One such choice is $\epsilon=\min_j\{\frac{1}{K_j}\}$, where $K_j=\sum_{i=1}^M A_{ij}$. In this case, θ_2 is given by

$$\begin{aligned}\theta_2^{(fs)} &= \frac{f_j(x_j^n)-f_j(\epsilon)}{x_j^n-\epsilon}\\ &= \sum_{i=1}^{M}\frac{A_{ij}^2 y_i}{p_i^{n2}-p_i^n A_{ij}(x_j^n-\epsilon)}\end{aligned}$$

and the update equation is still given by

$$x_j^{n+1} = \arg\min_{x_j \geq 0}\{F_j^{(fs)}(x_j) + P(x_j)\} \ .$$

We conjecture that this update strategy is globally convergent since it appears that the proof of [59] could be simply extended to handle this more general case.

3.4 Numerical Results

Our experimental results consist of trials using synthetic phantoms with functional values similar to human tissue in low dosage emission and transmission tomography simulations. The data are 128×128 projections, and the reconstruction is computed at a resolution of 128×128 pixels. We solve the MAP reconstructions consisting of two choices of p for the generalized Gaussian Markov random field (GGMRF) prior model with the prior log-density function of $P(x) = \sum_{i,j} \frac{b_{ij}}{p}\left(\frac{x_i - x_j}{\sigma}\right)^p$.

The experimental results of Figs. 3.1 and 3.2 show the convergence of the ICD/NR and ICD/FS for the emission case with a non-Gaussian prior, the emission case with a Gaussian prior, and the transmission case with a non-Gaussian prior. All trials employ sequential pixel updates in raster ordering. We note that with this particular non-Gaussian prior, the second derivative of the log-likelihood is not bounded, so the technical conditions of the proof are not met. However, we conjecture that the proof of [59] may also be generalized for this case.

The plots of Figures 3.1 and 3.2 experimentally show that the convergence properties of ICD/NR and ICD/FS are virtually identical. This is not surprising, since the log-likelihood function is generally close to quadratic and the values of $\theta_2^{(nr)}$ and $\theta_2^{(fs)}$ are therefore generally very close. While a proof for the global convergence of ICD/NR does not yet exist, its convergence appears consistently rapid.

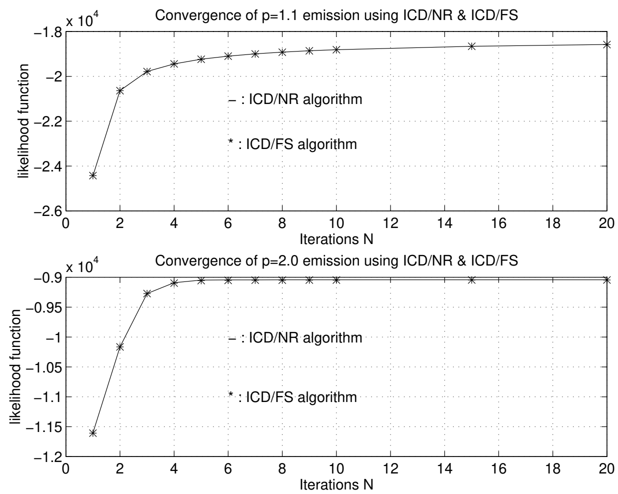


Fig. 3.1. Convergence of likelihood vs. iterations for the emission case with $p = 1.1$ and $p = 2.0$ GGMRF prior models, employing ICD/FS and ICD/NR algorithms.

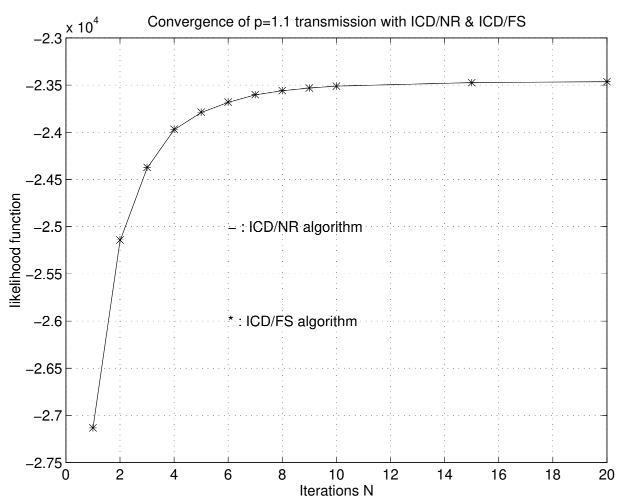


Fig. 3.2. Convergence of likelihood vs. iterations for the transmission case with $p = 1.1$ GGMRF prior model, employing ICD/FS and ICD/NR algorithms.

4. MULTIRESOLUTION NON-HOMOGENEOUS MRF MODEL

4.1 Introduction

The popularity of Bayesian methods in image processing applications such as image reconstruction [60] and restoration [8] is due to the ability of these methods to incorporate a model for the image in addition to the model for the data collection process. However, the performance of Bayesian methods is dependent on how accurately the model describes the underlying unknown image. Consequently, the field of image modeling has received considerable attention in the past decade.

Natural images often have edges and regions with different local characteristics. In some 3D problems, the material under study may also have structure known *a priori* to have behavior along one axis which differs from that along the other two [61]. It is therefore imperative for a good image model to adapt to such non-homogeneous and/or non-isotropic local behavior. One of the earliest attempts in this direction was by Geman and Geman when they proposed a doubly stochastic Markov random field (MRF) model with interacting discrete line processes to characterize the edges in the image [8].

However, there are several disadvantages associated with the above approach. First, it is difficult to model the discrete line processes. Second, the above formulation results in a difficult non-convex optimization problem, which requires computationally intensive stochastic relaxation techniques [46, 8] to achieve the approximate global minimum. While deterministic relaxation techniques [9, 52, 62] can be employed to reduce the computational cost, they often get trapped in a local minima. Finally, and perhaps most importantly, non-convexity is undesirable because the maximum *a* *posteriori* (MAP) reconstruction depends discontinuously on the data [12] resulting in unstable estimators [63, 64].

Alternatively, a host of continuously valued MRF models have been proposed [9, 2, 3, 10, 11, 12, 13] that provide good edge-preservation without explicitly modeling the discontinuities in an image. Most of these are distinguished by the choice of potential or regularization function that assigns cost to differences between neighboring pixels. This potential function can either be convex [3, 10, 12, 13] or non-convex [9, 2, 11]. However, convex potential functions have the advantage that they result in a tractable optimization problem. It is interesting to note that the use of an edge-preserving potential function in MRF models is equivalent to using non-interacting continuous valued line processes. This was first shown by Geman and Reynolds [11] by using a technique known as half-quadratic regularization [65, 66]. More recently, line processes obeying more complex smoothness constraints have also been proposed [67, 68].

The methods of anisotropic diffusion [69] have developed in parallel with Bayesian techniques, but have resulted in surprisingly analogous properties. As with continuously valued MRF's, anisotropic diffusion models employ a potential function that determines the space-varying diffusion coefficient. Edge-preservation then results by selectively diffusing along the edge rather than across it [70]. While anisotropic diffusion methods have been shown equivalent to optimization problems [70], they fundamentally differ from the Bayesian framework because the measured data enters the problem only as an initial condition.

Although MRF based models have been successfully used to improve the quality of the reconstructed image, present models have two important shortcomings: First, they only use very local information to model the behavior of the unknown image. Second, they are generally homogeneous with smooth regions, textures, and edges treated similarly. To overcome these shortcomings, a logical solution is to formulate a non-homogeneous MRF model with space-varying model parameters that adapts to the local characteristics of the image. The model could utilize large scale information to better model the local behavior in the vicinity of a pixel. However, such a formulation poses two difficulties: First, the number of model parameters to estimate is on the order of the number of pixels in the image. Second, the model parameters depend on the structure of the underlying image which is unknown. In some particular applications, *a priori* anatomical information may be used to guide the selection of these space-varying model parameters [71, 72]. However, in general there has been no effective method of estimating the very large number of parameters. Consequently, general non-homogeneous models have been largely ignored in the past.

In this Chapter, we formulate a novel non-homogeneous MRF model using a multiresolution framework [73, 74] that adapts to the local behavior of the underlying image. The basis of our development is the homogeneous generalized Gaussian MRF (GGMRF) model of Bouman and Sauer [12] that uses a single scale parameter to characterize the entire image. The reason for choosing the GGMRF rests on two important properties: First, the GGMRF has been shown to provide good edge-preservation [12]. Second, the scalable nature of GGMRF [13] yields a simple closed form ML estimate of the scale parameter [41]. The latter property motivates an intuitive local estimate of the space-varying scale parameter for the proposed non-homogeneous GGMRF. These space-varying scale parameters are analogous to continuous line processes. However, they differ from continuous line processes since they can take any value from zero to infinity. The purpose of the multiresolution framework is to enable the estimation of these space-varying local scale parameters at any resolution by using the image at the previous coarser resolution.

Using this non-homogeneous GGMRF image model and the sequential MAP (SMAP) estimator [75], we estimate the unknown image by starting at the coarsest resolution and proceeding recursively down to the finest resolution. An important advantage of such a formulation is that it results in a convex optimization problem at any resolution. Consequently, the reconstructed image at the finest resolution depends continuously on the measured data.

Since the local scale parameters estimated from the coarse resolution image may systematically over or under estimate the variation in the image, we introduce a global scaling parameter for the non-homogeneous GGMRF at each resolution. A useful property of this particular formulation is that the non-homogeneous GGMRF is still scalable with respect to this global scaling parameter. This property allows us to compute a closed form expression for the ML estimate of the global scale parameter. Since this estimate depends on the unknown image at each resolution, we use the EM algorithm [14, 15] to compute the ML estimate directly from the observed data during the reconstruction.

A desirable property of the proposed non-homogeneous GGMRF image model is that it automatically gives rise to multiresolution algorithms that estimate the image at progressively finer resolutions. Multiresolution algorithms have generated a lot of interest in image processing applications due to their faster convergence and better image modeling [76, 77, 78]. In particular, wavelet based methods have been proposed for tomographic applications and region-of-interest reconstruction [79, 80, 81]. Multiresolution methods are also particularly well suited to non-convex problems since they can more effectively avoid being trapped in some local minima than their single resolution counterparts [78].

Experimental results are presented for real transmission and emission data sets. These results indicate that our multiresolution algorithm substantially reduces computational cost while yielding better quality reconstructions as compared to the fixed resolution image models and reconstruction algorithms.

4.2 Multiresolution Non-Homogeneous MRF Model

In this section, after introducing the notation, we develop the multiresolution non-homogeneous MRF model starting from the homogeneous fixed resolution MRF model.

We use upper case letters to denote random quantities, while lower case letters denote the corresponding deterministic realizations of these random quantities. Let x

denote the unknown image that we are trying to model. The superscript k is used to denote the scale; $x^{(k)}$ therefore denotes the image at scale k. We use $k = 0$ to denote the finest resolution and $k = L$ to denote the coarsest resolution. At scale k, the set of lattice points is denoted by $S^{(k)}$, and the number of lattice points is denoted by N_k. We use a factor of 2 decimation in each dimension between two consecutive scales. Therefore the number of pixels reduce by a factor of 4 between two consecutive scales.

Typical past approaches have concentrated on the use of a homogeneous MRF's to model the image in the Bayesian framework. The log-likelihood of the image model that is generally used is of the form

$$\log P(x|\sigma) = - \sum_{\{i,j\}\in\mathcal{N}} b_{i-j}\rho(\frac{x_i - x_j}{\sigma}) + C \quad , \tag{4.1}$$

where $\mathcal{N}$ denotes the set of all neighboring pixel pairs, C is used to denote the normalization constant of the distribution that is independent of x, and $\rho(\cdot)$ is the potential function that assigns a cost to differences between neighboring pixels. The potential function is said to be scalable if

$$\rho(\frac{x}{\sigma}) = \frac{1}{\sigma^p}\rho(x) \quad ,$$

where $\sigma > 0$ and p is a constant.

The ease of optimal scale parameter estimation for the homogeneous MRF model (4.1) depends on the choice of the potential function $\rho(\cdot)$. It was shown in Chapter 2 that the ML estimate of the scale parameter can be computed in a computationally efficient manner for a host of potential functions. If the potential function is also scalable [12, 13], then the ML estimate of σ, given the unknown image, has a simple closed form expression. This is a useful property which we will exploit later on to obtain an intuitive estimate of the local scale parameter in the non-homogeneous model.

We use the generalized Gaussian MRF (GGMRF) of Bouman and Sauer [12] as the basis of our development for the non-homogeneous model because its scalable nature permits a simple expression for the ML estimate of σ. Furthermore, the GGMRF has been shown to provide good edge-preservation. The log-likelihood of the GGMRF is given as

$$\log P(x|\sigma) = -\frac{1}{p\sigma^p} \sum_{\{i,j\} \in \mathcal{N}} b_{i-j} \left| x_i - x_j \right|^p + C \quad . \tag{4.2}$$

The ML estimate of σ for the GGMRF is easy to compute (refer to Section 2.2.2) and is given as

$$\hat{\sigma}^p = \frac{1}{N} \sum_{\{i,j\} \in \mathcal{N}} b_{i-j} \left| x_i - x_j \right|^p \quad , \tag{4.3}$$

where N is the total number of pixels in the image x.

4.2.1 Non-homogeneous MRF model

Homogeneous MRF models (4.1) are in general unrealistic for natural images that have space-varying characteristics since a single scale parameter σ is used for the entire image. From this perspective, the logical solution is to replace the global scale parameter σ with a local scale parameter σ_{ij} that is different for each pair of neighboring pixels. Modifying the homogeneous GGMRF model (4.2) to have space-varying σ, we obtain

$$\log P(x|\{\sigma_{ij} | \forall (i,j) \in \mathcal{N}\}) = -\frac{1}{p} \sum_{\{i,j\} \in \mathcal{N}} b_{i-j} \left| \frac{x_i - x_j}{\sigma_{ij}} \right|^p + C \quad . \tag{4.4}$$

The advantage is that the above non-homogeneous GGMRF model can adapt to the local characteristics of the underlying image. However, the problem with the above formulation is that the number of scale parameters, $\{\sigma_{ij} | \forall \{i,j\} \in \mathcal{N}\}$, that have to be estimated is on the order of the number of pixels in the image.

To address this difficulty, we examine the form of the ML estimate of σ for the homogeneous case. Note that the global estimate of σ in (4.3) is obtained by averaging the local pixel differences over the entire image. Then intuitively the estimate of a local scale parameter at any pixel should only average the pixel differences in the neighborhood of the pixel. In the light of this observation, we define

$$\hat{\sigma}_i^p = \frac{1}{2} \sum_{j \in \mathcal{N}_i} b_{i-j} \; \left| x_i - x_j \right|^p \quad , \tag{4.5}$$

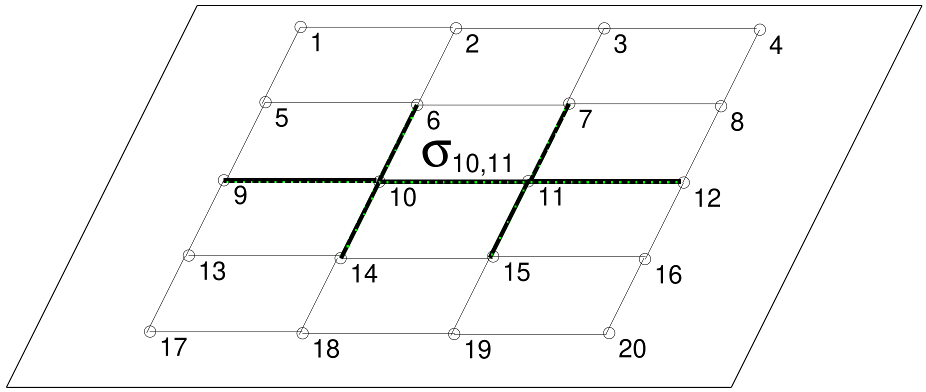


Fig. 4.1. The above figure shows the pixel differences that are averaged to obtain the estimate of $\sigma_{10,11}$ when a 4-point neighborhood is used.

where σ_i is now the local scale parameter at pixel i and $\mathcal{N}_i$ denotes the neighborhood of the pixel i. Note that the factor of 2 in (4.5) is to ensure that the average of local scale parameters is equal to the global scaling parameter when x is homogeneous, i.e $\hat{\sigma}^p = \frac{1}{N}\sum_i \hat{\sigma}_i^p$. The scale parameter for the link between pixels i and j, σ_{ij}, is then estimated as the local average of σ_i and σ_j.

$$\begin{aligned} \hat{\sigma}_{ij}^p &= \frac{1}{2}\left\{\hat{\sigma}_i^p + \hat{\sigma}_j^p\right\} \\ &= \frac{1}{4}\left\{\sum_{q\in\mathcal{N}_i} b_{i-q}\,|x_i - x_q|^p + \sum_{q\in\mathcal{N}_j} b_{j-q}\,|x_j - x_q|^p\right\} \quad . \end{aligned} \tag{4.6}$$

Figure 4.1 illustrates the pixel differences that are averaged to obtain a particular estimate of the scale parameter between pixels 10 and 11 for a four-point neighborhood.

4.2.2 Multiresolution framework

The problem with the formulation of the non-homogeneous GGMRF model (4.4) is that the estimate of the local scale parameter σ_{ij} depends on the unknown image x. The limitation of this approach is that random samples of x are not available to estimate σ_{ij}. Moreover, ML estimation of σ_{ij} from the data y is not desirable because the number of unknown parameters is approximately equal to the number of observations in y. Consequently, the ML estimates of σ_{ij} are likely to be poorly behaved and pose numerical difficulties.

To address this problem, we use a multiresolution framework as shown in Fig. 4.2. The unknown image is represented at multiple scales, $x = \{x^{(0)}, \ldots, x^{(L)}\}$, and the local scale parameters at any resolution k are estimated from the image at the previous resolution $k+1$.

Let I_{k+1}^{k} denote the linear interpolation operator from scale $k+1$ to scale k. Let

$$z^{(k)} = I_{k+1}^{k} x^{(k+1)} \tag{4.7}$$

denote the interpolated image at scale k. Then the local scale parameter estimate between pixels i and j given by (4.6) is modified to depend on the interpolated image $z^{(k)}$ instead of the image x as follows

$$\hat{\sigma}_{ij}^{p}(x^{(k+1)}) = \frac{1}{4}\left\{ \sum_{q\in\mathcal{N}_i} b_{i-q}\left|z_i^{(k)} - z_q^{(k)}\right|^p + \sum_{q\in\mathcal{N}_j} b_{j-q}\left|z_j^{(k)} - z_q^{(k)}\right|^p \right\} \ . \tag{4.8}$$

Figure 4.2 shows the dependence of the local scale parameters on the interpolated image from the previous resolution.

Using the multiresolution framework and the modified expression for $\hat{\sigma}_{ij}$ given by (4.8), the non-homogeneous GGMRF model (4.4) is written as

$$\log P(x^{(k)}|x^{(k+1)}) = -\frac{1}{p}\sum_{\{i,j\}\in\mathcal{N}} b_{i-j}\left|\frac{x_i^{(k)} - x_j^{(k)}}{\hat{\sigma}_{ij}(x^{(k+1)})}\right|^p + C \quad k \neq L \ . \tag{4.9}$$

Note that the dependence of image $x^{(k)}$ on image $x^{(k+1)}$ is only through $\hat{\sigma}_{ij}(x^{(k+1)})$. The above model is only valid for scales $k \neq L$ since at the coarsest resolution L we do not have an interpolated image from which the local scale parameters may be estimated. Instead we let $X^{(L)}$ be a homogeneous GGMRF with the log-likelihood given in (4.2). This is justified because in the Bayesian framework the amount of data increases significantly at coarser resolutions thereby decreasing the importance of the image model at these resolutions.

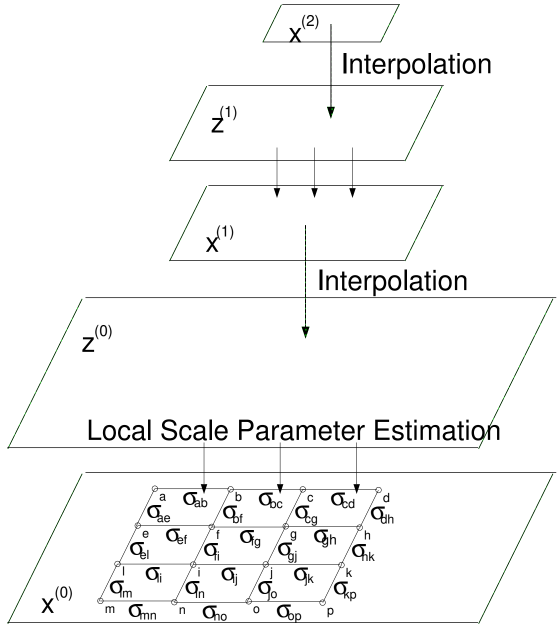


Fig. 4.2. The above figure shows the multiresolution framework for the non-homogeneous GGMRF image model. The local scale parameters at any resolution are estimated using the interpolated image from the previous resolution.

4.2.3 Global scaling parameter

One problem that may arise by estimating σ_{ij} using (4.8) from the interpolated image is that it may over or under estimate the image variation by a fixed constant. To address this, we introduce another parameter λ_k that acts as a global scaling parameter for the entire image at resolution k. The modified conditional log-likelihood is then given as

$$\log P(x^{(k)}|x^{(k+1)}, \lambda_k) = \frac{-1}{p\lambda_k^p} \sum_{\{i,j\}\in\mathcal{N}} b_{i-j} \left| \frac{x_i^{(k)} - x_j^{(k)}}{\hat{\sigma}_{ij}(x^{(k+1)})} \right|^p + C \qquad k \neq L \quad (4.10)$$

$$\log P(x^{(L)}|\lambda_L) = \frac{-1}{p\lambda_L^p} \sum_{\{i,j\}\in\mathcal{N}} b_{i-j} \left| x_i^{(L)} - x_j^{(L)} \right|^p + C \ . \quad (4.11)$$

Since $X^{(k)}$ only depends on $X^{(k+1)}$, the collection of random fields $X = \{X^{(0)}, \ldots, X^{(L)}\}$ form a first order Markov chain in scale. The log-likelihood of the complete image model is then easily obtained as

$$\log P(x|\lambda) = \sum_{k=0}^{L-1} \log P(x^{(k)}|x^{(k+1)}, \lambda_k) + \log P(x^{(L)}|\lambda_L) \quad . \tag{4.12}$$

We will use x to denote the collection of images, $\{x^{(0)}, \ldots, x^{(L)}\}$, for the rest of this chapter.

4.3 Multiresolution Reconstruction Algorithm

In this section we present the multiresolution reconstruction algorithm by specifying the estimation criterion for the unknown image. We also develop the multiresolution data model for the tomography problem. Figure 4.3 summarizes the complete tomographic image reconstruction algorithm developed in this section.

Let y denote the measurements and let $P(y|x)$ denote the probability of observing y given the image x. A popular criterion for estimation of the unknown image is the maximum *a posteriori* (MAP) estimate given as

$$\begin{aligned} \hat{x} &= \arg\max_x \{\log P(y|x) + \log P(x|\lambda)\} \\ &= \arg\max_x \left\{ \log P(y|x^{(0)}) + \sum_{k=0}^{L-1} \log P(x^{(k)}|x^{(k+1)}, \lambda_k) + \log P(x^{(L)}|\lambda_L) \right\} . \end{aligned}$$

The MAP estimate in this case is not desirable since it poses a formidable optimization problem. More importantly, the MAP estimate is not suitable in the multiresolution framework since it assigns equal weights to errors at all the different resolutions.

On the other hand, the sequential MAP (SMAP) estimator of Bouman and Shapiro [75] is perhaps more suitable in the multiresolution framework since the cost criterion that it approximately minimizes assigns higher weights to errors at the coarser resolution. Moreover, the SMAP estimator is desirable because it results in a tractable optimization problem.

The SMAP estimation procedure starts at the coarsest scale L and proceeds down to the finer scales as follows

$$\hat{x}^{(L)} = \arg\max_{x^{(L)}} \left\{ \log P(y|x^{(L)}) + \log P(x^{(L)}|\lambda_L) \right\} \tag{4.13}$$

$$\hat{x}^{(k)} = \arg\max_{x^{(k)}} \left\{ \log P(y|x^{(k)}) + \log P(x^{(k)}|\hat{x}^{(k+1)}, \lambda_k) \right\} \quad k = L-1, \ldots, 0 \tag{4.14}$$

Note that the procedure is recursive since the reconstructed image $\hat{x}^{(k+1)}$ at scale $k+1$ is used to compute the reconstruction $\hat{x}^{(k)}$ at scale k.

4.3.1 Data models for tomography

We specify the multiresolution data model $\log P(y|x^{(k)})$ for tomographic problems in this section. The models for emission and transmission tomography are the same as described in Section 2.3.1. However, we reintroduce the notation in this Section explicitly accounting for the scale at which the unknown image x is represented.

Let $x^{(k)}$ denote the column vector of emission intensities in the emission case or the attenuation densities in the transmission case at scale k. Let y denote the column vector of measurements of Poisson-distributed photon counts at the detectors for all angles and displacements. Note that the data is collected at a fixed resolution.

For the emission case, let A_{ij} be the probability that a photon emitted from cell $j \in S^{(0)}$ is registered at the i^{th} detector. Let A be the projection matrix with elements $\{A_{ij}\}$, and let A_{i*} denote the i^{th} row of the projection matrix. Then, for the emission case, the log-likelihood of the data given the image at scale 0 is

$$(emission) \quad \log P(y|x^{(0)}) = \sum_{i=1}^{M} \left(-A_{i*}x^{(0)} + y_i \log\{A_{i*}x^{(0)}\} - \log(y_i!)\right) \ , \tag{4.15}$$

where M is the total number of projections.

The transmission case is similar, but has A_{ij} corresponding to the length of intersection between the j^{th} ($j \in S^{(0)}$) cell and the i^{th} projection. Let the dosage photon counts be Poisson-distributed with rate y_T. Then, for the transmission case, the log-likelihood of the data given the image at scale 0 is

$$(transmission) \quad \log P(y|x^{(0)}) = \sum_{i=1}^{M} \left(-y_T e^{-A_{i*}x^{(0)}} + y_i(\log y_T - A_{i*}x^{(0)}) - \log(y_i!)\right) \ . \tag{4.16}$$

Note that (4.15) and (4.16) are similar in form and can be compactly written as

$$\log P(y|x^{(0)}) = -\sum_{i=1}^{M} f(y_i, A_{i*}x^{(0)}) \ , \tag{4.17}$$

where

$$f(y_i, A_{i*}x^{(0)}) = A_{i*}x^{(0)} - y_i \log\{A_{i*}x^{(0)}\} + \log(y_i!) \quad (emission) \tag{4.18}$$

$$= y_T e^{-A_{i*}x^{(0)}} - y_i(\log y_T - A_{i*}x^{(0)}) + \log(y_i!) \quad (trans.). \tag{4.19}$$

Notice that for both the emission and transmission cases $f(y_i, \cdot)$ is a convex function. Henceforth, in the interest of notational simplicity, we will use (4.17) as the log-likelihood for emission as well as transmission tomography.

For scales $k > 0$, the likelihood of the data can be computed using Bayes rule as follows

$$P(y|x^{(k)}) = \int P(y|x^{(k-1)})P(x^{(k-1)}|x^{(k)})dx^{(k-1)} \quad . \tag{4.20}$$

However, the above integral is intractable to compute analytically for the proposed image model. Therefore, exclusively for the purpose of computing the likelihood of the data at scale k, we approximate the image model at scales $n < k$ as follows

$$\mathcal{P}(X^{(n)} = x^{(n)}|x^{(n+1)}) = \begin{cases} 1 & \text{if } x^{(n)} = \tilde{I}_{n+1}^{n} x^{(n+1)} \\ 0 & \text{otherwise} \end{cases} \quad , \tag{4.21}$$

while retaining the original model given by (4.10) and (4.11) for scales $n \geq k$. Note that $\tilde{I}_{n+1}^{n}$ denotes an interpolation scheme different from the one in (4.7). The modification of the image model for scales $n < k$ is an approximation since we cannot legitimately change our image model during the estimation process. However the approximation is justified since the data likelihood function has only a secondary dependence on the image model and it primarily depends on the observations. This is reminiscent of the hybrid graph structure used in [75].

Substituting the modified image model given by (4.21) into (4.20) and using (4.17), we obtain the log-likehood of the data at scale k as

$$\begin{aligned} \log P(y|x^{(k)}) &= -\sum_{i=1}^{M} f(y_i, A_{i*} \prod_{i=0}^{k-1} \tilde{I}_{i+1}^{i} x^{(k)}) \\ &= -\sum_{i=1}^{M} f(y_i, A_{i*}^{(k)} x^{(k)}) \quad , \end{aligned}$$

where we define

$$A^{(k)} = A \prod_{i=0}^{k-1} \tilde{I}_{i+1}^{i} = A\tilde{I}_k^0 \ . \tag{4.22}$$

Note that $A^{(k)}$ now takes on the role of the projection matrix at scale k.

4.3.2 Decimation of data

Note that in the previous development, the full resolution data is retained as we move to coarser resolutions of the image. This results in a large amount of data to estimate relatively few pixel values at the coarser resolutions. Therefore considerable computational savings can be realized by decimating the data also without seriously affecting the performance of the reconstruction algorithm at coarser resolutions. Note that we use decimation here to denote the combined filtering and down-sampling operation. However, if a Poisson data model is used, the computational savings by decimating the data can only be realised for the emission case. This is because the rate of the Poisson data counts is linearly related to the projections of the image in the emission case while they have a non-linear relationship in the transmission case. On the other hand, if a quadratic approximation [50] is used for the Poisson data model or no filtering of the data is done prior to down-sampling, then computational savings can also be realised in the transmission case. However, simply down-sampling the data without filtering is not desirable since it could significantly affect the performance of the algorithm at coarser resolutions. With this in view and the fact that we use only the Poisson data model in this Chapter, we will restrict ourselves to decimating the data in the emission case only and retain the full resolution data for the transmission case.

Let $y^{(l)}$ denote the decimated data at scale l obtained from the data at the finest resolution y as follows

$$y^{(l)} = \tilde{I}_0^l y \ ,$$

where $\tilde{I}_0^l$ is the decimation operator that decimates the sinogram image y from scale 0 to scale l. In particular, we use

$$\tilde{I}_0^l = \left(\tilde{I}_l^0\right)^T \quad , \tag{4.23}$$

where ^{T} denotes the transpose operation. Note that since a factor of 4 is omitted in (4.23), $\tilde{I}_0^l$ results in summation rather than an average over blocks of the sinogram image y. For the emission case, since the elements of Y are Poisson with rates $Ax^{(0)}$, it follows that the elements of $Y^{(l)}$ are Poisson with rates $\tilde{I}_0^l Ax^{(0)}$. Let M_l be the number of projections in $y^{(l)}$. Then using (4.21), (4.20), and (4.15), the log-likelihood of $y^{(l)}$ given $x^{(k)}$ can be written as

$$\log P(y^{(l)}|x^{(k)}) = \sum_{i=1}^{M_l} \left(-A_{i*}^{(lk)} x^{(k)} + y_i^{(l)} \log\{A_{i*}^{(lk)} x^{(k)}\} - \log(y_i^{(l)}!)\right) \quad , \tag{4.24}$$

where now the projections matrix is given as

$$A^{(lk)} = \tilde{I}_0^l A \tilde{I}_k^0 \quad . \tag{4.25}$$

We now use $P(y^{(l)}|x^{(k)})$ instead of $P(y|x^{(k)})$ in (4.13) and (4.14) to compute the SMAP estimate. These final set of equations are shown in Fig. 4.3. We use the iterative coordinate descent (ICD) algorithm [50] with Newton-Raphson updates to compute the minimization required for the SMAP estimate. The ICD method works for both the emission and the transmission case. However, we are not restricted to using ICD and other optimization methods for tomographic problems in the Bayesian framework can also be employed [4, 2, 3, 82].

As we move to coarser resolutions, the data can be decimated along with the image, i.e $l = k$. However, since the decimation of data results in some loss of information, it might be desirable to stagger the decimation of the data with respect to that of the image. In practice we found that staggering the decimation of the data by one level, i.e using $l = [k-1]_+$ ($[x]_+ \triangleq \max\{x, 0\}$), yields results that are almost indistinguishable from those obtained by no decimation of the data, but provides computational savings.

For $k = L, \ldots, 0$ {

1. set $l = [k-1]_+$ for emission case, set $l = 0$ for transmission case.
2. Generate if necessary $y^{(l)} = \tilde{I}_0^l y$ and $A^{(lk)} = \tilde{I}_0^l A \tilde{I}_k^0$.
3. Compute $\hat{\lambda}_k$ using (4.29) or (4.30).
4. If $k = L$ then use ICD to compute the SMAP estimate

$$\hat{x}^{(L)} = \arg\min_{x^{(L)}} \left\{ \sum_{i=1}^{M_l} f(y_i^{(l)}, A_{i*}^{(lL)} x^{(L)}) + \frac{1}{p\hat{\lambda}_L^p} \sum_{\{i,j\}\in\mathcal{N}} b_{i-j} \left| x_i^{(L)} - x_j^{(L)} \right|^p \right\}$$

else

$$\hat{x}^{(k)} = \arg\min_{x^{(k)}} \left\{ \sum_{i=1}^{M_l} f(y_i^{(l)}, A_{i*}^{(lk)} x^{(k)}) + \frac{1}{p\hat{\lambda}_k^p} \sum_{\{i,j\}\in\mathcal{N}} \frac{b_{i-j}}{\hat{\sigma}_{ij}^p(\hat{x}^{(k+1)})} \left| x_i^{(k)} - x_j^{(k)} \right|^p \right\}$$

where $f(\cdot,\cdot)$ is defined in (4.18) and (4.19) for emission and transmission tomography.

}

Fig. 4.3. Complete image reconstruction algorithm for emission and transmission tomography.

In order to quantify the computational gains obtained by decimating the data, we define the weights

$$w_k = \frac{T_k}{T_0} \ , \tag{4.26}$$

where T_k is the CPU time taken to compute one full iteration of the ICD algorithm at scale k. Table 4.1 shows the weights computed for an emission phantom. The listed weights show the computational benefits obtained by decimating the data at coarser resolutions as compared to using the full resolution data. Table 4.2 shows the weights for the transmission case. Note that in this case no data decimation is used.

Table 4.1 The computed weights $w_k = T_k/T_0$ are shown for the emission case when ICD with Newton-Raphson updates is used to perform the optimization. The 3 cases shown are: 1) data is not decimated with image, 2) data is decimated along with image, 3) data decimation is staggered by one level with respect to the image.

data decimation	w_0	w_1	w_2	w_3	w_4	$\sum_i^4 w_i$
No ($l = 0$)	1.000	0.310	0.107	0.042	0.0176	1.477
Yes ($l = k$)	1.000	0.186	0.038	0.009	0.0019	1.235
Staggered ($l = [k-1]_+$)	1.000	0.310	0.054	0.012	0.0021	1.378

Table 4.2 The computed weights $w_k = T_k/T_0$ are shown for the transmission case when the full resolution data is retained at the coarser resolutions. ICD with Newton-Raphson updates is used to perform the optimization.

data decimation	w_0	w_1	w_2	w_3	w_4	$\sum_i^4 w_i$
No ($l = 0$)	1.000	0.351	0.139	0.061	0.029	1.580

4.3.3 Choice of the interpolation scheme

In this section, we discuss the choice of linear interpolation schemes for operators $\tilde{I}_{k+1}^k$ and I_{k+1}^k.

The interpolation scheme used in the modified image model (4.21) is used to compute the projection matrix at scale k using $A^{(k)} = A\tilde{I}_k^0$ or $A^{(lk)} = \tilde{I}_0^l A\tilde{I}_k^0$ depending on whether the data is decimated or not. Therefore a complex interpolation scheme with a large support would tend to make $A^{(k)}$ or $A^{(lk)}$ less sparse rendering the optimization at image scale k to be computationally expensive. With this in view, we choose $\tilde{I}_{k+1}^k$ to be simple pixel replication. Another advantage of this choice is that now the elements of $A^{(k)}$ can be computed by using the same algorithm that was used to compute the elements of A. The only difference is that $A^{(k)}$ is computed using pixels that are 2^k times larger in each dimension than the pixels at scale 0. This method is computationally cheaper than computing the equivalent matrix multiplication $A\tilde{I}_k^0$.

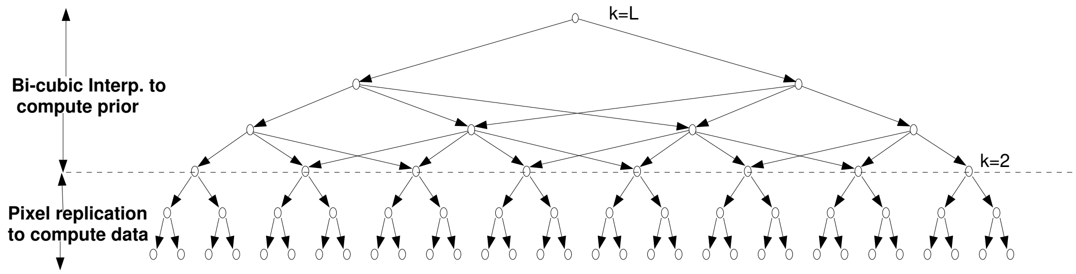


Fig. 4.4. The above figure shows the different interpolation schemes used to compute the data model and the prior model when computing the reconstruction at scale $k = 2$. Pixel replication is used for scales $k < 2$ and bi-cubic interpolation is used for scales $k \geq 2$.

On the other hand, a complex interpolation scheme to compute $z^{(k)} = I_{k+1}^{k} x^{(k+1)}$ does not affect the computational complexity of the optimization at image scale k. Moreover, a superior interpolation scheme is required to obtain a good estimate of the local scale parameters σ_{ij} for scale k. We therefore choose bi-cubic interpolation for this purpose [83]. Figure 4.4 shows the hybrid graph structure used for the image model when computing the reconstruction at scale $k = 2$.

4.3.4 ML estimation of the global scaling parameter

For the proposed scheme to be unsupervised, we need to estimate the global scaling parameters λ_k directly from the data y before computing the SMAP estimate of the image at each scale k. We use the conditional ML estimate for λ_k defined as

$$\hat{\lambda}_L = \arg\max_{\lambda_L} P(y^{(l)}|\lambda_L) \tag{4.27}$$

$$\hat{\lambda}_k = \arg\max_{\lambda_k} P(y^{(l)}|\hat{x}^{(k+1)}, \lambda_k) \quad k \neq L \ , \tag{4.28}$$

where the scale l for the data depends on the particular decimation scheme used for the data. Note that for scales $k < L$, the ML estimate is conditioned on the SMAP estimate $\hat{x}^{(k+1)}$ at the previous scale, which is available when ML estimation of λ_k is performed at scale k. Since the conditional ML estimate of λ_k depends on the data $y^{(l)}$ only through the unknown image $x^{(k)}$, this is a classic example of the incomplete data problem. We use the EM algorithm [14, 15] that was specifically designed to solve problems of this nature to obtain the estimates. Using the result of Appendix B, the iterative update equation for the conditional ML estimate of λ_L is obtained as

$$\hat{\lambda}_L^p(t+1) = \frac{1}{N_L} \sum_{\{i,j\}\in\mathcal{N}} b_{i-j} E\left[\left|X_i^{(L)} - X_j^{(L)}\right|^p | y^{(l)}, \hat{\lambda}_L(t)\right] \tag{4.29}$$

and the set of update equations for the conditional ML estimate of λ_k, $k \neq L$, is obtained as

$$\hat{\lambda}_k^p(t+1) = \frac{1}{N_k} \sum_{\{i,j\}\in\mathcal{N}} \frac{b_{i-j}}{\hat{\sigma}_{ij}^p(\hat{x}^{(k+1)})} E\left[\left|X_i^{(k)} - X_j^{(k)}\right|^p | y^{(l)}, \hat{x}^{(k+1)}, \hat{\lambda}_k(t)\right] \quad , \tag{4.30}$$

where t denotes the discrete iteration index. The expectation involved in the updates is intractable to compute analytically. Instead we numerically compute the expectation by using samples generated from the posterior distribution $P(x^{(L)}|y^{(l)}, \hat{\lambda}_L(t))$ or $P(x^{(k)}|y^{(l)}, \hat{x}^{(k+1)}, \hat{\lambda}_k(t))$. The fast simulation technique described in Section 2.3.3 is used for this purpose. It was shown in Section 2.4.2 that using a single sample to compute the expectation yields approximately the same parameter at convergence as using multiple samples to compute the expectation. We therefore use only one sample to compute the expectation in the update equations. Also, rather than summing the local pixel differences in (4.29) and (4.30) over the entire image, we restrict it to the support of the object to be reconstructed. This is to ensure that the reconstructions are not overly smooth due to the inclusion of the constant zero background. Details of the computation of the support of the object are given in Section 2.4.2.

4.4 Applications to Optical Flow

Optical flow is the distribution of apparent velocities of movement of brightness patterns in an image. The problem of determining the optical flow from a sequence of intensity images is truly an ill-posed one. This is because the velocity field at each image point has two components while the change in image brightness at a point in the image plane due to motion yields only one constraint. In fact, as it turns out, we can only recover the component of the velocity field along the gradient of the brightness at any image point [84]. To address this difficulty, traditional methods incorporate an additional constraint of smoothness of the velocity field while computing the optical flow. This is equivalent to employing an image model for the velocity field. In particular, MRF based models for the motion vector field in the Bayesian framework have been proposed [85, 86, 87]. The smoothness constraint works well for motion vectors within a single object but obviously fails at object boundaries where there is a sharp discontinuity in the motion field. Consequently edge-preserving velocity field models find useful application for motion estimation.

4.4.1 Data model and optimization strategy for optical flow

In this section, we develop a 3 frame multiresolution data model for computing the optical flow. Let $y(t)$ be the intensity image at time instance t. We assume that $y(t)$ is corrupted by i. i. d. Gaussian noise with variance σ_n^2. Let the data y denote the set of intensity images, $\{y(0), y(1), y(2)\}$, at 3 time instances namely $t = 0, 1, 2$. Let $x^{(k)}$ denote the field of vector displacements that represents the optical flow at scale k for time $t = 1$. Let

$$E_f(s, x^{(0)}) \triangleq y_{s+x_s^{(0)}}(2) - y_s(1) \tag{4.31}$$

be the forward prediction error and let

$$E_b(s, x^{(0)}) \triangleq y_{s-x_s^{(0)}}(0) - y_s(1) \tag{4.32}$$

be the backward prediction error. Since the data y is available only at discrete spatial positions $s \in S$, we require an interpolation scheme to compute $y_{s+x_s^{(0)}}(2)$ and $y_{s-x_s^{(0)}}(0)$ in (4.31) and (4.32) respectively. We will use bi-cubic interpolation for this purpose [83].

For image points that are neither revealed at $t = 1$ nor become occluded at $t = 2$, the backward prediction error or the forward prediction error can be minimized to compute the optical flow. However, for image points that are revealed at time $t = 1$, the backward prediction error is inappropriate to minimize since the revealed points are missing in $y(0)$. Similarly, for points that become occluded at time $t = 2$, the forward prediction error is inappropriate to minimize since the occluded points are missing in $y(2)$. Therefore to account for the occluded and revealed points in the image, the data model should switch between the forward and backward prediction errors in a space-varying fashion. The difficulty is that we do not know *a priori* which points in the image are going to be revealed or occluded. However, note that the absolute backward prediction error would in general be lower than the absolute forward prediction error for an occluded point and vice versa for a revealed point. With this observation and the fact that the noise is Gaussian distributed with variance σ_n^2, we formulate the log-likelihood of the data given $x^{(0)}$ as

$$\log P(y|x^{(0)}) = -\frac{1}{2\sigma_n^2} \sum_{s \in S^{(0)}} \min(E_f^2(s, x^{(0)}), E_b^2(s, x^{(0)})) \ .$$

The likelihood $P(y|x^{(k)})$ given by (4.20) is again difficult to compute exactly. As discussed in Sect. 4.3.1, $\log P(y|x^{(k)})$ can be approximated by

$$\log P(y|x^{(k)}) = -\frac{1}{2\sigma_n^2} \sum_{s \in S^{(0)}} \min(E_f^2(s, \tilde{I}_k^0 x^{(k)}), E_b^2(s, \tilde{I}_k^0 x^{(k)})) \ .$$

The image model that we use for the optical flow field is the same as the non-homogeneous model described in Sect. 4.2. The only difference is that $x_s^{(k)}$ is now a vector with two components $x_s^{(k)}(1)$ and $x_s^{(k)}(2)$ corresponding to the displacement of the pixel at spatial position s in two orthogonal directions. The $|\cdot|^p$ function in all the equations in Sect. 4.2 is now interpreted as

$$\begin{aligned} |v|^p &\triangleq |v(1)|^p + |v(2)|^p \\ &= ||v||_p^p \ , \end{aligned}$$

where v is any vector with two components $v(1)$ and $v(2)$.

We use the SMAP estimation procedure described in Sect. 4.3 to compute the optical flow. ICD is again used to perform the optimization required in (4.13) and (4.14). However, we restrict $x_s^{(k)}$ to take discrete values in the set

$$R = \{-5, -4.75, -4.5, \ldots, 4.5, 4.75, 5\} \times \{-5, -4.75, -4.5, \ldots, 4.5, 4.75, 5\}.$$

The maximum is then obtained by brute force evaluation of the SMAP cost criterion for all $x_s^{(k)} \in R$.

4.5 Experimental Results

4.5.1 Tomography

In this section, we experimentally study the convergence speed and reconstruction quality of the proposed multiresolution non-homogeneous GGMRF model and the multiresolution reconstruction algorithm. We use synthetic as well as real data sets for both emission and transmission tomography.

The transmission data that we use was acquired by measuring projections of a real flashlight. The original data consisted of 1024 projections taken at 792 equally spaced angles. From this original data, we generated a low resolution data set by retaining every 4^{th} projection at every 4^{th} angle. We used the full resolution data to form a "ground truth" image using convolution back projection (CBP) reconstruction shown in Fig. 4.7(a). All other reconstructions with this data set were then done using the lower resolution data. This approach allows us to obtain a pseudo ground truth image, which is normally not available for real data sets.

The 3D SPECT data that we use was acquired from cardiac perfusion imaging using Tc-99m sestamibi for a male patient. For each slice, 128 projections were taken at 120 uniformly spaced angles between 0 and 2π. Figure 4.9(a) shows the CBP reconstruction of one of the slices. The reconstruction was done at 128 by 128 0.356cm pixel resolution. The total photon count for this slice was 148761.

Figure 4.10(a) shows the synthetic head phantom that we use for emission tomography. Figure 4.10(b) shows the corresponding CBP reconstruction. The emission rates are on an array of 128 by 128 pixels of size 1.56mm^2. The data was generated by taking 128 Poisson distributed projections at each of 128 uniformly spaced angles between 0 and π. The total photon count was approximately 3 million.

Since the proposed multiresolution non-homogeneous model is derived from the fixed resolution GGMRF model, the performance of the new model is compared to that of the fixed resolution GGMRF model. We refer to the proposed multiresolution method using the non-homogeneous GGMRF as the MR method and the fixed resolution method using the homogeneous GGMRF as the FR method. We use the iterative coordinate descent (ICD) algorithm with Newton-Raphson updates in both the cases to perform the optimization. Eight different update ordering schemes are used to facilitate the convergence of the algorithm [6]. A fixed value of 1.1 is used for p in all the reconstructions. For the emission case, we stagger the decimation of the data by one level with respect to the image in the MR method. For the transmission case, the data is not decimated in the MR method.

The coarsest resolution L for the MR method is determined by decimating the full resolution image at scale 0 till we obtain a grid of 8x8 pixels. Therefore, L is set to 5 for the flashlight data and set to 4 for the 3D SPECT data and the synthetic emission data. For the MR method, ICD is initialized at the coarsest resolution $k = L$ with a constant image set to the mean obtained from all the projections. For the FR method, ICD is initialized using the CBP image at the finest resolution $k = 0$.

Let $N[k]$ be the total number of full iterations at scale k performed by ICD in the MR method. Then the equivalent number of iterations at base resolution for the MR method is defined as

$$N_{\text{eq}} = \sum_{k=0}^{L} w_k \, N[k] \quad , \tag{4.33}$$

where w_k is the relative cost of computation between scale k and scale 0 defined in Sec. 4.3.2. For the FR method, $N_{\text{eq}} = N[0]$. We use N_{eq} when comparing the computation of the FR and the MR method since it represents equivalent amount of computational effort in both the cases.

For the purpose of estimating the global scale parameters in both the FR and the MR method, we restricted the computation of the estimates to the support of the object. For the MR method, a decimated version of the support of the object is used at the coarser resolution. We use the ML estimate of the global scaling parameters in all our reconstructions for both the MR as well as the FR method.

In order to experimentally study the speed of convergence, we require a meaningful measure of convergence. The posterior log-likelihoods are not directly comparable since they pertain to two different models. A more meaningful measure is the normalized root mean square error (NRMSE). NRMSE for the FR method is defined as

$$\mathrm{NRMSE}(n) = \left(\frac{(x_n - x_c)^T (x_n - x_c)}{x_c^T x_c} \right)^{1/2} ,$$

where x_n is the image at iteration n of ICD and x_c is obtained by performing 500 iterations of ICD. We set the global scaling parameter λ to the ML estimate for generating both x_c and x_n. NRMSE for the MR method is defined as

$$\mathrm{NRMSE}(n) = \left(\frac{(x_n^{(0)} - x_c^{(0)})^T (x_n^{(0)} - x_c^{(0)})}{{x_c^{(0)}}^T x_c^{(0)}} \right)^{1/2} .$$

In this case 500 iterations of ICD is performed at all the resolutions starting from $k = L$ and proceeding down to $k = 0$ to obtain $x_c^{(0)}$ for fixed values of λ_k. $x_n^{(0)}$ denotes the image at scale 0 for n equivalent iterations at scale 0, i.e $N_{\mathrm{eq}} = n$. We set the global scaling parameters λ_k to the ML estimates for generating both $x_c^{(0)}$ and $x_n^{(0)}$.

We have to specify the number of iterations performed at each resolution with respect to the iterations at the base resolution in the MR method. Since the computational cost of a single iteration decreases at coarser resolutions, it makes sense to perform more iterations at these resolutions. This also yields faster convergence since effects produced by local updates are propagated globally in a more efficient manner at coarser resolutions. The particular rule that we use to determine the number of iterations at scale k is given as

$$N[k] = \left\lceil 2^{k/a} N[0] \right\rceil ,$$

where $\lceil b \rceil$ denotes the smallest integer greater than or equal to b and $a > 0$ controls the growth of the number of iterations at coarser resolutions. We found experimentally that when a is small, $N[k]$ grows rapidly yielding better convergence for small $N[0]$. However, when $N[0]$ is large, a small a increases the computation and does not improve the convergence significantly. Alternatively, a large a results in poor convergence when $N[0]$ is small. Generally, we found that $a = 3$ performed well over a large range of $N[0]$. Therefore, we will use $N[k] = \left\lceil 2^{k/3} N[0] \right\rceil$ to compute the number of iterations at scale k for the MR method.

Figure 4.5 shows the NRMSE of the MR and the FR method for the flashlight data with respect to N_{eq}. In this case we notice that the FR method has lower NRMSE for $N_{\mathrm{eq}} < 13$ iterations. However, the MR method has a faster rate of convergence as seen from the slope of the plots and becomes better than the FR method after 13 iterations. Figures 4.6(a) and 4.6(b) show the NRMSE of the MR and the FR method for the 3D SPECT data and the synthetic emission data respectively. For the 3D SPECT data, the MR method is slightly better than the FR method in terms of NRMSE for the same amount of computation. For the synthetic emission data, the MR method is better than the FR method for $N_{\mathrm{eq}} > 3$.

Figure 4.7 shows the reconstructions corresponding to the flashlight data. Figure 4.7(c) and Figure 4.7(d) show the reconstruction obtained by the FR method and the MR method respectively. Both the FR and the MR method use ML estimates for the global scaling parameter λ. Notice that the MR method suppresses noise while retaining the details and edges more effectively that the FR method. This is more obvious in Figure 4.8 that shows blown up versions of the images in Figure 4.7. Figure 4.7(e) shows the local scale parameters estimated at the finest resolution in the MR method.

Figure 4.9 shows the reconstructions corresponding to a single slice of a male patient for the 3D SPECT data set. Note in this case we do not have any "ground truth" to verify the accuracy of the reconstructions. The MR methods again yields a cleaner and sharper reconstruction than the FR method. Figure 4.10 shows the reconstructions for the synthetic emission data. In this case the original image in Fig. 4.10(a) can be used to qualitatively assess the performance of the FR and the MR method. From Figs. 4.10(c) and 4.10(d) it is clear that the MR method is suppressing the noise in the smooth regions while retaining the sharp edges in a much better fashion than the FR method. It is possible to reduce the noise in the smooth regions in Fig. 4.10(c) for the FR method by using a larger scaling parameter λ.

However, in doing so we also blur the edges and the detail in the image. The MR method does not suffer from this because the local scale parameters vary according to the image characteristic selectively smoothing in the smooth regions and retaining detail where there are sharp edges.

Figure 4.11 shows the ML estimates obtained for the global scale parameters λ_k at resolutions $k = 0, \ldots, L-1$, using the EM algorithm. In order to speed up the convergence, the EM algorithm at each resolution was initialized with the ML estimate obtained at the previous resolution. Figure 4.12(a) and Figure 4.12(b) show the corresponding plots for the 3D SPECT data and the synthetic emission data. Note that the general trend is that the global scaling parameters decrease with increasing resolution making the reconstructions smoother and more regularized.

4.5.2 Optical flow

In this section, we present experimental results on optical flow using synthetic as well as real data to compare the performance between the homogeneous GGMRF model and the multiresolution non-homogeneous GGMRF model. Since the data model for optical flow is highly non-convex, fixed resolution methods usually tend to get stuck in a local minimum. However, the non-homogeneous GGMRF model is less prone to this problem since it is implemented in a multiresolution framework. Therefore to make a fair comparison between the two models, we compute the reconstruction corresponding to the homogeneous GGMRF model in a multiresolution framework as well. The only difference is that a single scale parameter is used in the homogeneous case while the scale parameter is space-varying in the non-homogeneous case. The interpolated image from the previous coarser resolution is used in both cases to initialize the optimization at the next finer resolution.

Figure 4.13(a) shows the first and the third frame of a synthetic image sequence used to compute the optical flow. Two dimensional separable sine waves with additive white Gaussian noise are used to generate the images. The object is moving one pixel per frame diagonally upwards to the left and the background is moving 1.5 pixels per frame diagonally downwards to the right. Figure 4.13(b) shows the true optical flow for this sequence.

Figure 4.13(c) shows the motion vectors estimated using the homogeneous GGMRF with $p = 1.0$ at the finest resolution. Figure 4.13(d) shows the corresponding estimates using the non-homogeneous GGMRF with $p = 1.0$. The variance of the additive Gaussian noise (σ_n^2) is assumed to be unknown in both cases and is directly estimated from the data along with the scale parameters for the image model. The reconstructions show clearly that the non-homogeneous field model preserves the discontinuity in the flow field in a much better fashion than the homogeneous field model.

Figures 4.14(a-c) show frames 13, 14, and 15 from the salesman sequence used to compute the optical flow. Figure 4.14(d) shows the average frame difference between frames 13 and 14, and frames 14 and 15. The difference image highlights the areas where motion is taking place. Figures 4.15(a-b) show the motion vector estimates obtained at scale $k = 2$ for the salesman sequence for the non-homogeneous GGMRF model ($p = 1.0$) and the homogeneous GGMRF model ($p = 1.0$) respectively. The motion vectors are not scaled and represent the actual displacement. Again, the variance of the noise, σ_n^2, was assumed unknown and was estimated directly from the images. Notice that the motion vectors for the left thumb of the salesman is more accurately estimated in the non-homogeneous case than the homogeneous case as seen from Fig. 4.15(a) and Fig. 4.15(b). The non-homogeneous model also yields a cleaner estimate in the background where the motion vectors are zero than the homogeneous model. Similar observations can be made from Fig. 4.16 and Fig. 4.17 which show the motion vectors estimates for both the non-homogeneous case and the homogeneous case at scale $k = 1$ and $k = 0$ respectively. Note that the motion vectors in Fig. 4.16 and Fig. 4.17 are scaled to fit the grid and do not represent the actual displacements.

4.6 Conclusion

We have formulated a novel non-homogeneous MRF image model using a multiresolution framework. The model is capable of adapting to the local behavior of the underlying image by employing space-varying scale parameters that are estimated from the coarser resolution image. This variation in parameters allows edges, textures and smooth regions of the image to be modeled differently. Moreover, the proposed non-homogeneous MRF model and the multiresolution reconstruction algorithm for emission and transmission tomography results in tractable convex optimization problems at all the resolutions. Consequently, the reconstructed image at the finest resolution depends continuously on the data. We have demonstrated experimentally that the multiresolution non-homogeneous GGMRF model is superior to the fixed resolution homogeneous GGMRF model in terms of quality of reconstruction and slightly better in terms of computational efficiency.

The multiresolution framework, in addition to facilitating the estimation of the local scale parameters, also yields a computationally efficient and robust algorithm that estimates the unknown image at progressively finer resolutions in a recursive manner. Consequently, other complex problems with highly non-linear data models such as geophysical structure estimation from seismic data, inverse optical tomography, and optical flow computation from a sequence of image could conceivably benefit from the proposed algorithm since it is less likely to be trapped in a local minima.

Furthermore, the proposed non-homogeneous/multiresolution reconstruction strategy is quite general since other more complex parameters can be estimated similarly. For example, parameters can be estimated that represent directional properties of local edges and textures. Perhaps spatially non-homogeneous values of the shape parameter p can be estimated using the method suggested by Jeff and Pun [39].

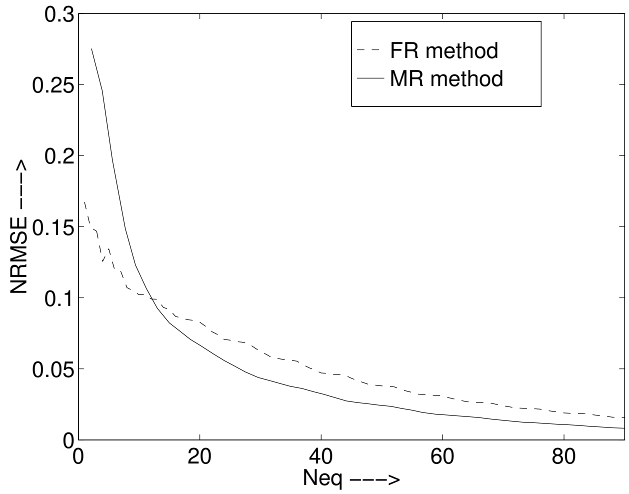


Fig. 4.5. Comparison of convergence rates for the FR and the MR method for the flashlight data. The FR method is initialized using the CBP image while the MR method is initialized at the coarsest resolution ($k = 5$) using the mean obtained from the projections. The NRMSE is plotted with respect to the equivalent iterations at scale 0 (N_{eq}) so that the computational cost is comparable between the FR and the MR method.

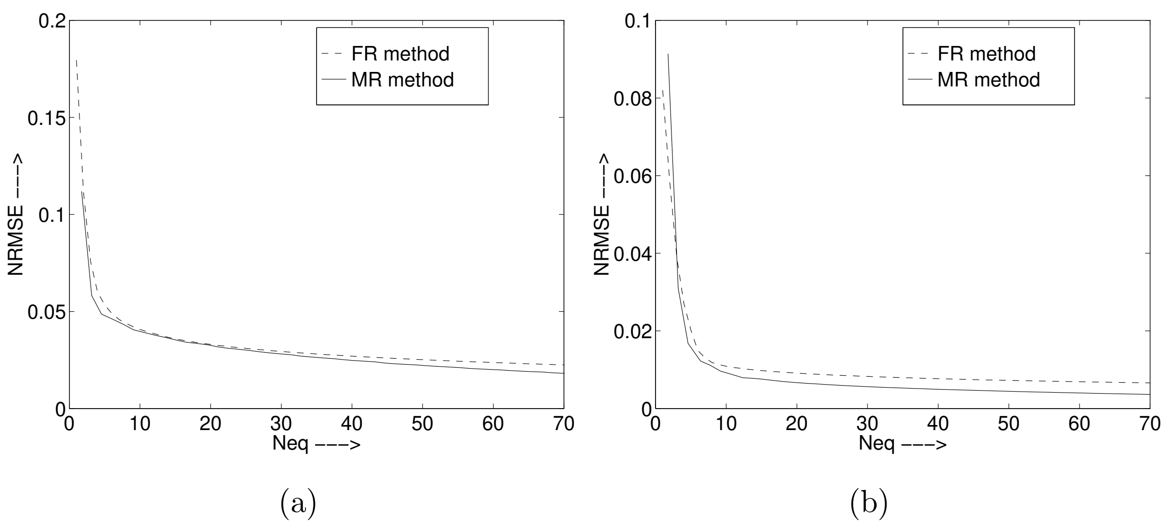
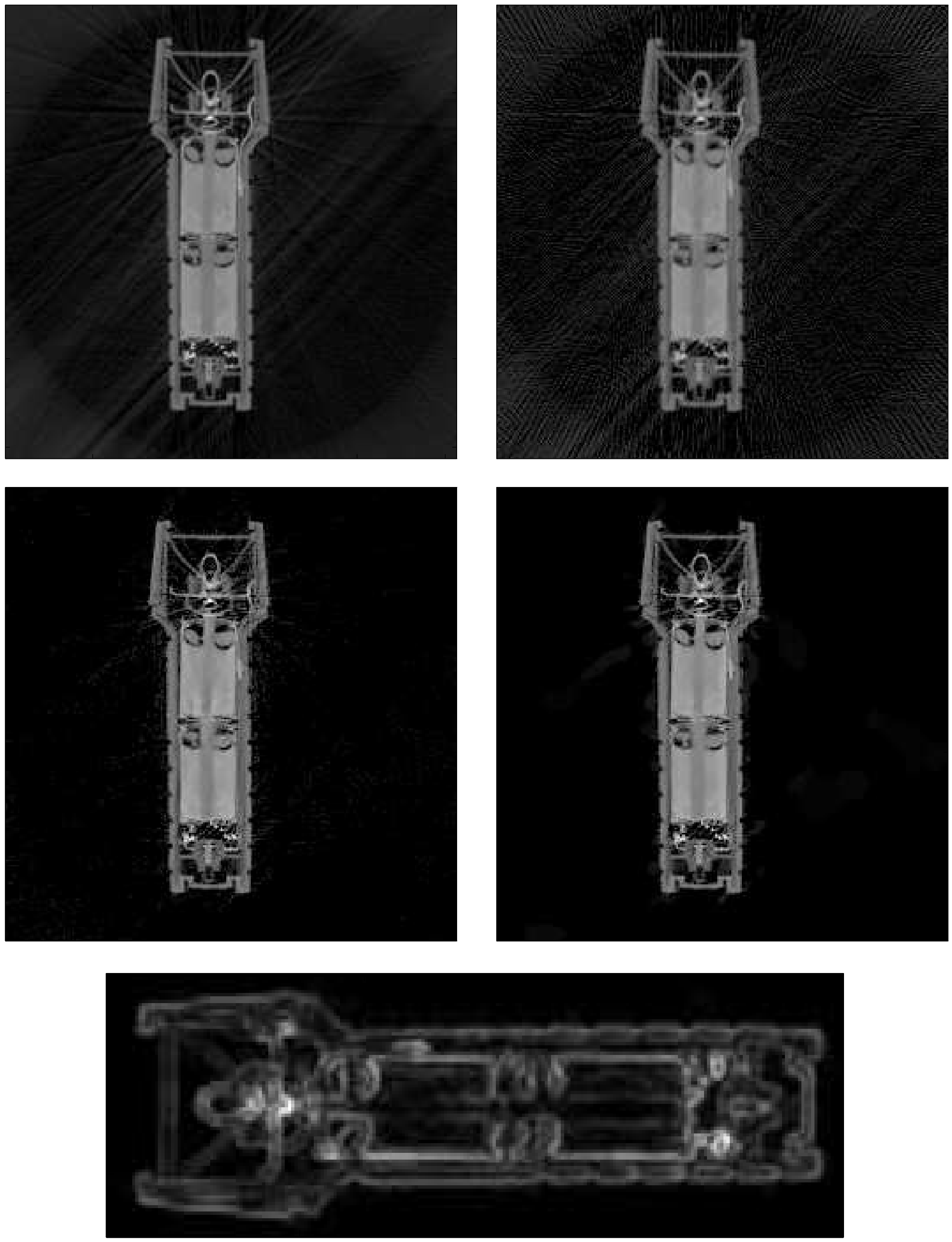


Fig. 4.6. Comparison of convergence rates for the FR and the MR method for the (a) 3D SPECT data and (b) synthetic emission data. The FR method is initialized using the CBP image while the MR method is initialized at the coarsest resolution ($k = 4$) using the mean obtained from the projections. The NRMSE is plotted with respect to the equivalent iterations at scale 0 (N_{eq}) so that the computational cost is comparable between the FR and the MR method.

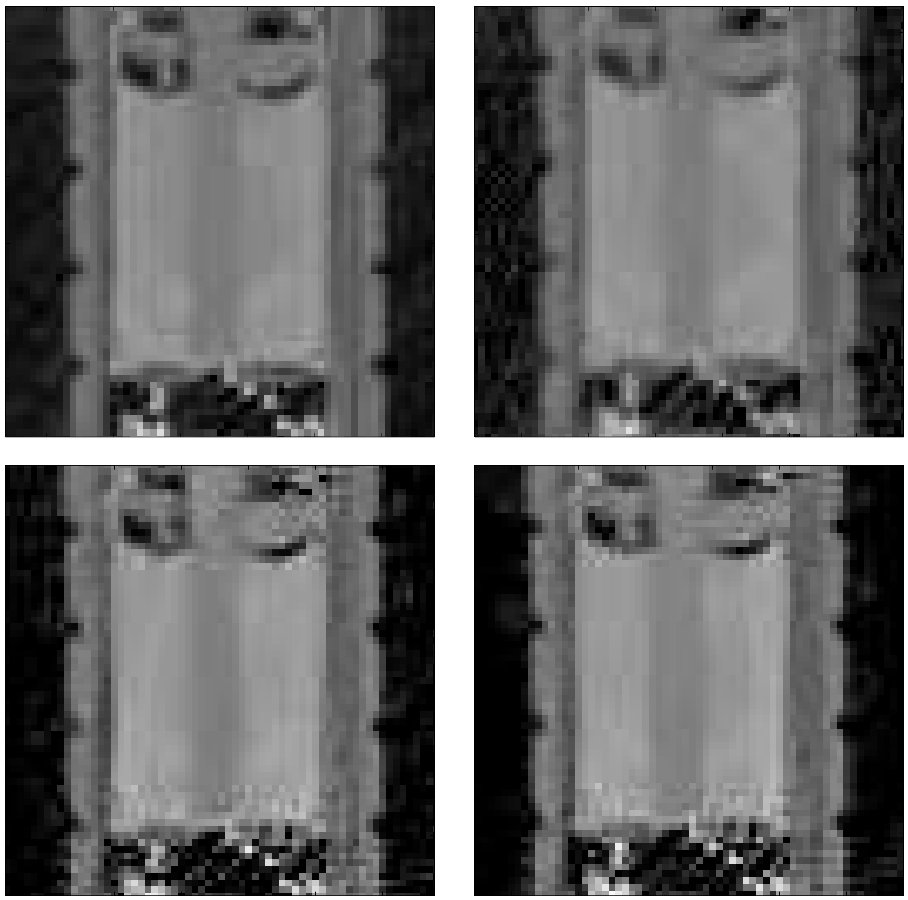


(a) (b)

(c) (d)

(e)

Fig. 4.7. Images corresponding to real transmission data of a flashlight. (a) "Ground truth", (b) CBP image, reconstructions using (c) FR method with ML estimate for global scaling parameter, (d) MR method with ML estimates of global scaling parameters at all the resolutions, and (e) Local scale parameters estimated at the finest resolution for the MR method. (Data courtesy of Trent Neel, Wright-Patterson Air Force Base & Nicolas Dussausoy, Aracor.)



(a) (b) (c) (d)

Fig. 4.8. Blow-up of images corresponding to Fig. 4.7. (a) “Ground truth”, (b) CBP image, reconstructions using (c) FR method, and (d) MR method.

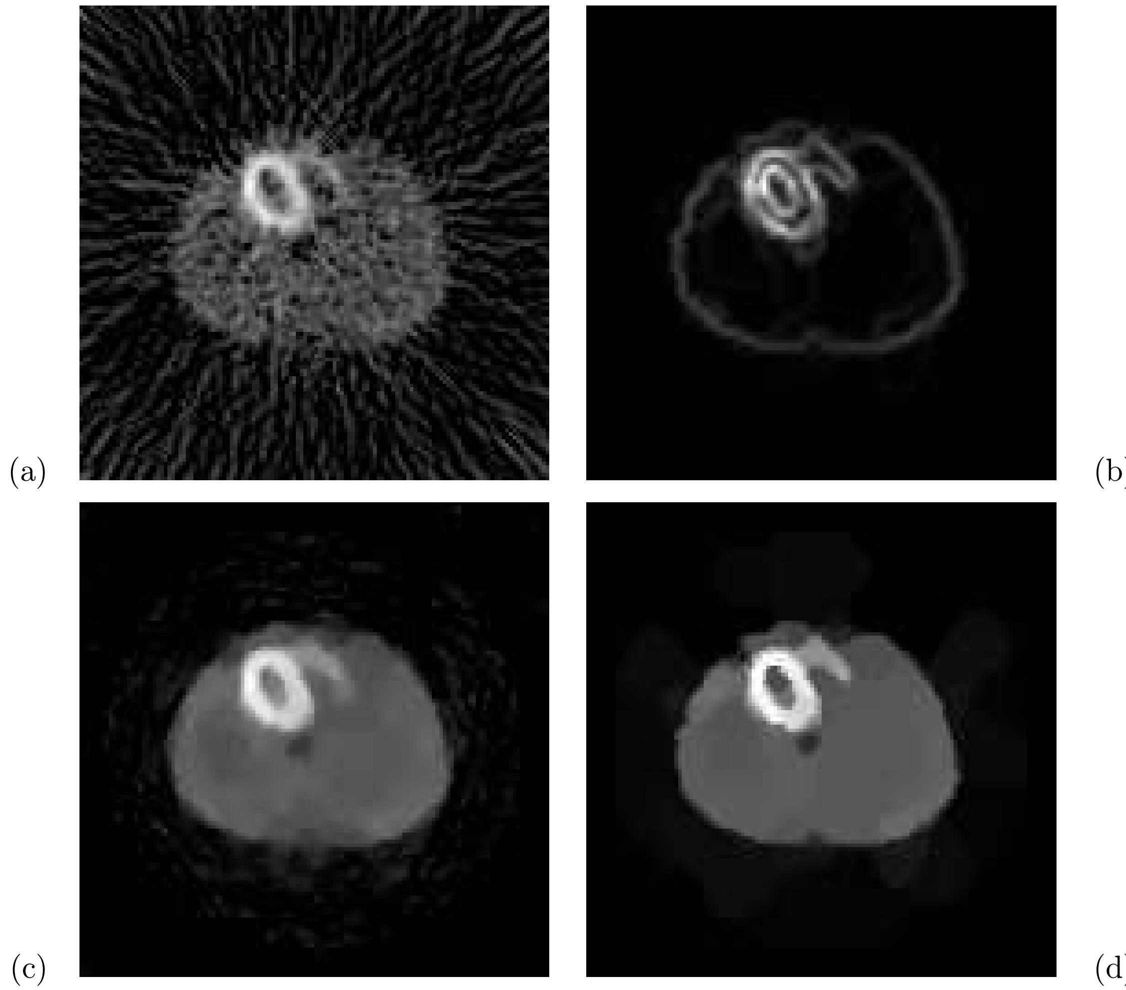
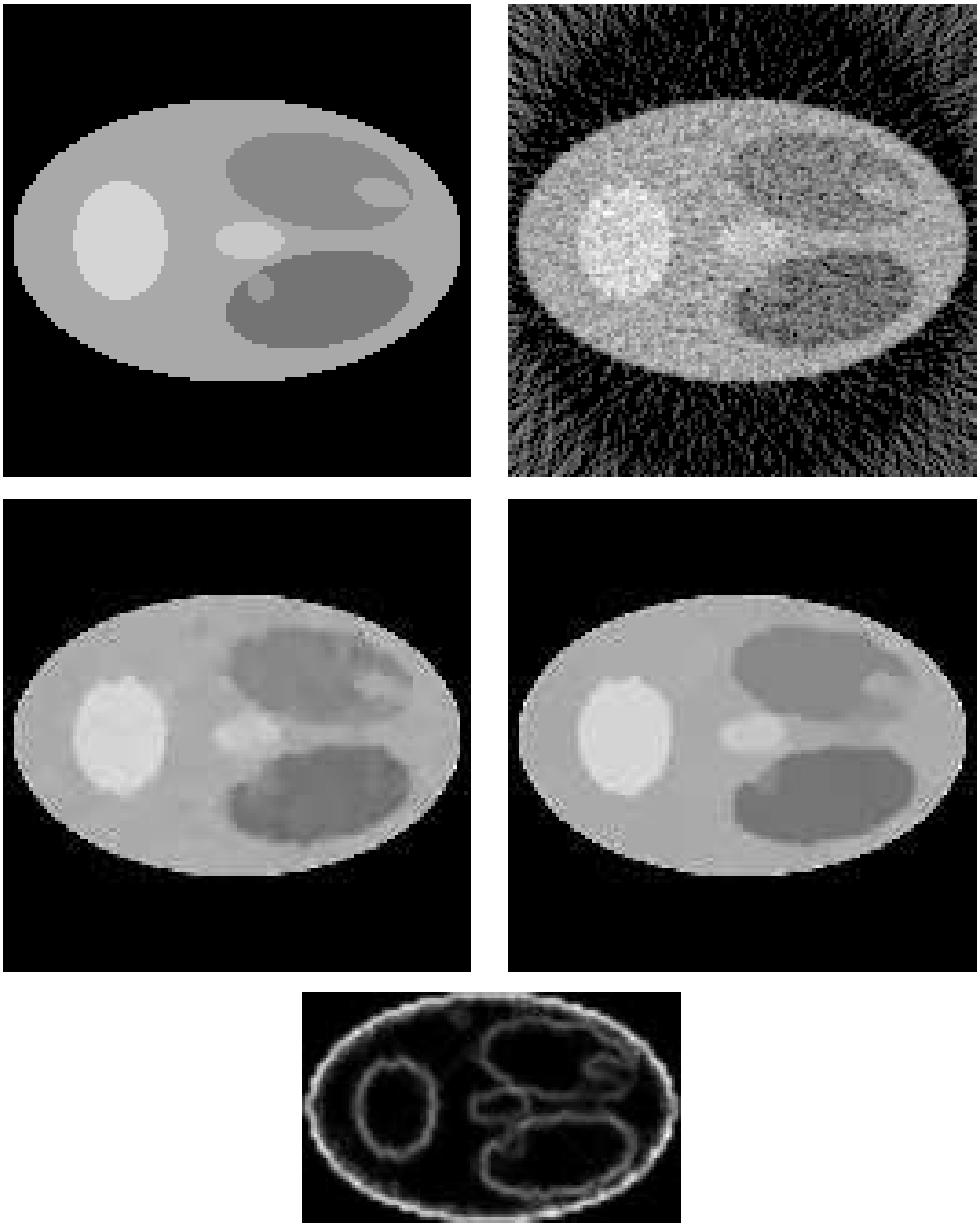


Fig. 4.9. Images corresponding to a single slice of the real 3D SPECT data for cardiac perfusion imaging for a male patient. (a) CBP, (b) Local scale parameters estimated at the finest resolution for the MR method, reconstructions using (c) FR method with ML estimate for scale parameter, (d) MR method with ML estimates of global scaling parameters at all the resolutions. (Data courtesy of Tin-Su Pan & Michael A. King of the University of Massachusetts.)



(a) (b) (c) (d) (e)

Fig. 4.10. Images corresponding to synthetic emission data of a head phantom. (a) Original, (b) CBP image, reconstructions using (c) FR method with ML estimate for global scaling parameter, (d) MR method with ML estimates of global scaling parameters at all the resolutions, (e) Local scale parameters estimated at the finest resolution for the MR method.

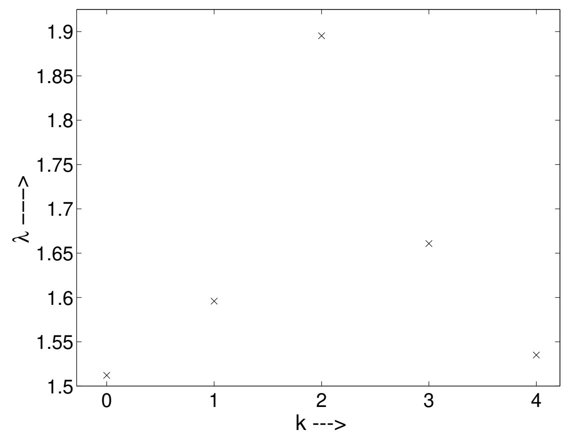
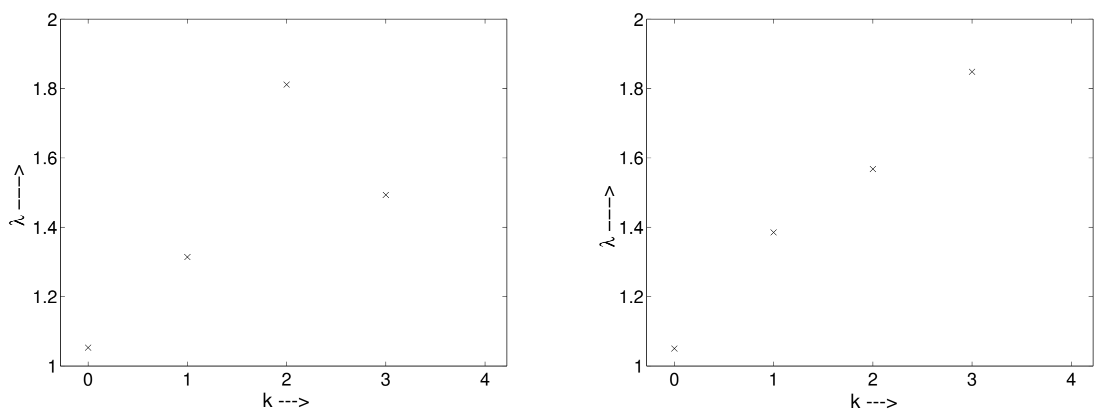


Fig. 4.11. Global scaling parameter estimates at resolutions $k = 0, \ldots, L-1$, obtained using the EM algorithm for the flashlight data. At any resolution, the EM algorithm was initialized using the ML estimate of the global scaling parameter at the previous resolution.



(a) (b)

Fig. 4.12. Global scaling parameter estimates at resolutions $k = 0, \ldots, L-1$, obtained using the EM algorithm for (a) 3D SPECT data and (b) synthetic emission data. At any resolution, the EM algorithm was initialized by using the ML estimate of the global scaling parameter at the previous resolution.

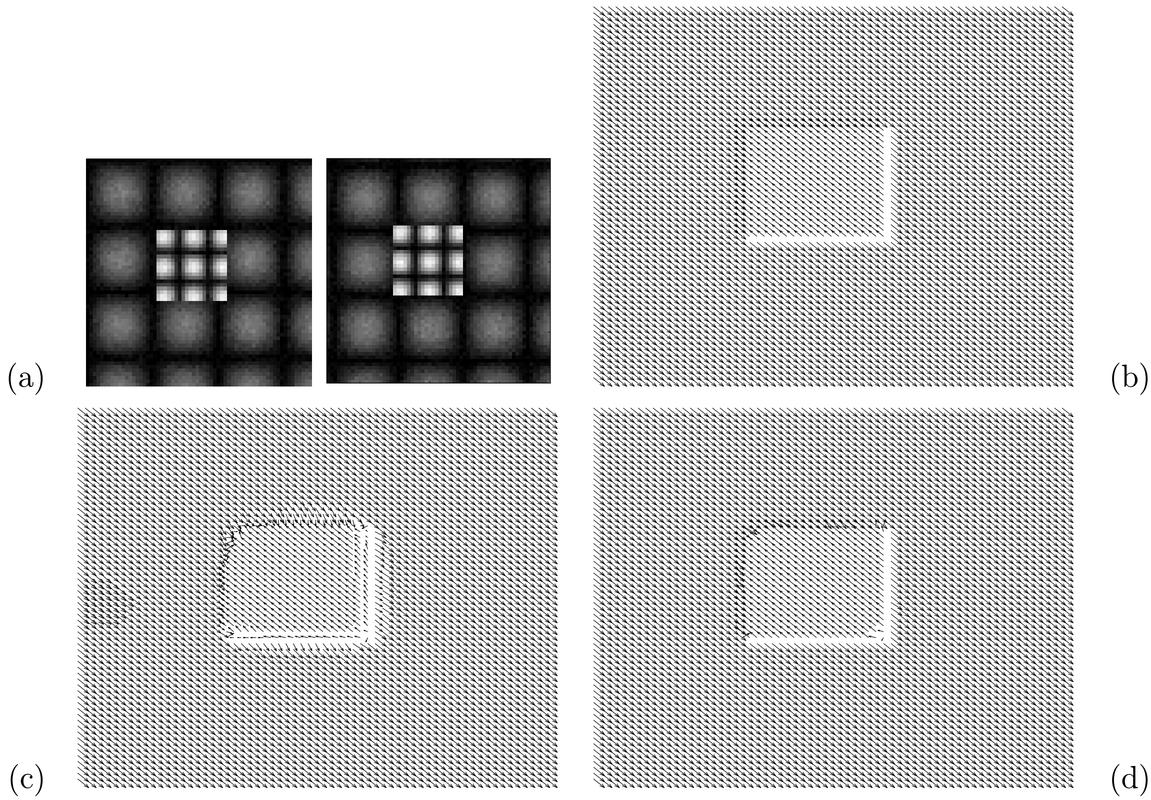


Fig. 4.13. (a) Two consecutive frames from the synthetic data sequence corrupted by Gaussian noise. (b) True motion vectors. The object has moved 1 pixel diagonally upwards and the background has moved 1.5 pixels diagonally downwards. Estimated motion vectors at scale $k = 0$ using the (c) homogeneous GGMRF model and the (d) non-homogeneous GGMRF model with $p = 1.0$. Note both estimates are obtained in a multiresolution framework where the coarse resolution image is used as the initial condition for the next finer resolution. The noise variance σ_n^2 is estimated directly from the data in both cases.

(a) (b)

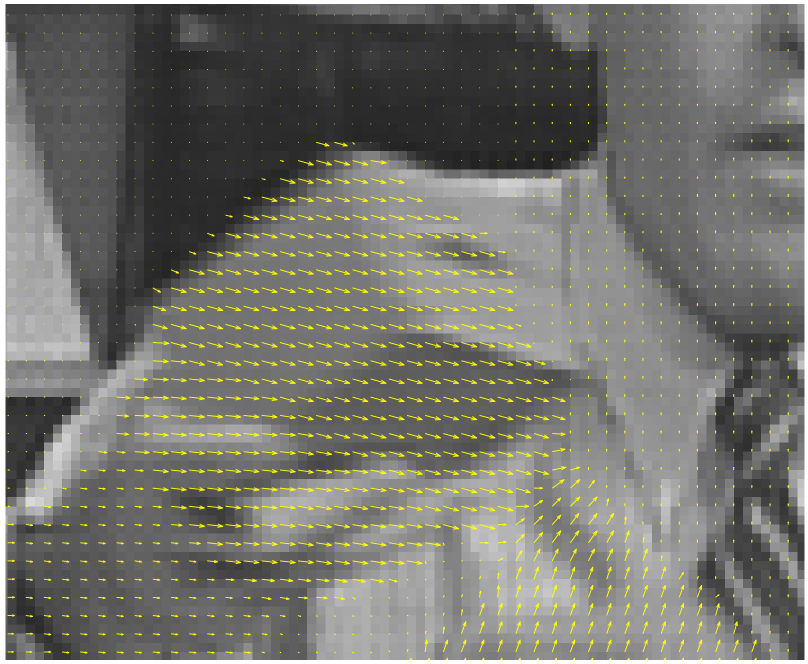
(c) (d)

Fig. 4.14. Frames (a) 13, (b) 14, and (c) 15 of the salesman sequence. (d) Shows the average absolute difference between frames 13 and 14 and frames 14 and 15.

(a)

(b)

Fig. 4.15. Comparison of estimated motion vectors at scale $k = 2$ using the (a) non-homogeneous GGMRF model and the (b) homogeneous GGMRF model with $p = 1.0$. Note both estimates are obtained in a multiresolution framework where the coarse resolution image is used as the initial condition for the next finer resolution. The noise variance σ_n^2 is estimated directly from the data in both cases.



(a)

(b)

Fig. 4.16. Comparison of estimated motion vectors at scale $k = 1$ using the (a) non-homogeneous GGMRF model and the (b) homogeneous GGMRF model with $p = 1.0$. Note both estimates are obtained in a multiresolution framework where the coarse resolution image is used as the initial condition for the next finer resolution. The noise variance σ_n^2 is estimated directly from the data in both cases.

(a)

(b)

Fig. 4.17. Comparison of estimated motion vectors at scale $k = 0$ using the (a) non-homogeneous GGMRF model and the (b) homogeneous GGMRF model with $p = 1.0$. Note both estimates are obtained in a multiresolution framework where the coarse resolution image is used as the initial condition for the next finer resolution. The noise variance σ_n^2 is estimated directly from the data in both cases.

5. IMAGE RECONSTRUCTION FROM TIME-RESOLVED DIFFUSION DATA

5.1 Introduction

Medical optical tomography has generated considerable interest in recent times [88]. The advantage of using infrared light as an imaging modality stems from the fact that it is non-ionizing and hence can be used for continuous monitoring. More importantly, it is capable of revealing the functioning of the tissues as opposed to other standard imaging modalities that reflect only the static physical structure. A final advantage is that optical imaging systems can be made portable, making them useful in clinical situations such as surgery, trauma, and intensive care.

An accurate model for the propagation of photons through tissue can be obtained from transport theory. There are two basic approaches using this theory: an essentially discrete model of individual photon interactions, such as Monte-Carlo [89], or a continuous model based on a differential equation approximation, such as the diffusion equation. While the Monte-Carlo method is more generally applicable, it is computationally expensive to implement. On the other hand, the diffusion approximation is accurate for highly scattering media (which is the case for tissues) while being computationally tractable. Therefore we will use the diffusion equation as our data model.

The inverse problem of reconstructing the absorption and scattering coefficients from diffuse measurements of light is highly nonlinear. To facilitate the computation of the unknown coefficients, several approaches attempt to locally linearize the original inverse problem. The linear perturbation model [90, 91, 92] is one such method that employs a Taylor series approximation about a reference distribution for the unknown coefficients to obtain a set of linear equations. The associated weight matrix is computed using Monte-Carlo simulations [93]. However, the utility of this method is limited if we do not have a reference distribution that is close to the actual distribution of the unknown coefficients. Alternatively, the Newton-Raphson (NR) method employs a Taylor series expansion about the current estimate of the unknown coefficients to obtain a more refined estimate. This procedure is then iterated until convergence. The NR method has also been used with the Levenberg-Marquardt procedure to control the nonlinearity of the underlying problem [94]. Other numerical optimization methods that have been used include projection onto convex sets (POCS) [90], simultaneous algebraic reconstruction technique (SART) [92], SART-type algorithm, [95] and conjugate gradient descent (CGD) [93]. We will be using a modified version of CGD for our optimization because of its superior convergence properties. The modification involves replacing simple line searching with bent-line searching to enforce the positivity constraint on the unknown coefficients.

The inverse problem is known to be ill-posed and some form of regularization is necessary to make the solution well behaved. Toward this end, we formulate the inverse problem in the Bayesian framework and use the maximum *a posteriori* (MAP) estimation criterion to compute the reconstruction. This approach enables us to incorporate *a priori* knowledge of the unknown field through an image model to regularize the solution. The regularization methods that have previously been employed include truncated singular-value decomposition, Tikhonov and constraint-based regularization that impose a penalty on the norm of the solution vector [94]. However, none of these methods properly model edges that are normally present in the unknown field for real objects, resulting in smooth reconstructions with blurred edges. An important contribution of the present work is the use of an edge-preserving generalized Gaussian Markov random field (GGMRF) model [12] for the unknown field.

To compute the likelihood of the data, we need to solve the diffusion equation forward in time. The finite element method (FEM) has been widely used for this purpose [94, 96, 97]. However, we propose to do this by discretizing the diffusion equation using a finite-difference approach [98]. The discretization can be accomplished in a number of ways since the spatial derivatives can be evaluated at the present (implicit) or the past (explicit) time instance. In particular, we use an alternating-directions method [99] that computes the spatial derivative implicitly for one spatial direction and explicitly for the other spatial direction in the first half of the time step. In the next half time step, the implicit and explicit directions are switched. This method is known to be stable even for large time steps. Also, by virtue of the alternating-directions, the resulting matrix that needs to be inverted in the forward computation is tridiagonal. The inversion can therefore easily be done in $O(N)$ time, where N is the number of pixels in the unknown image.

Since we will use CGD to compute the MAP estimate, we require the gradient of data likelihood with respect to the unknown field. The novelty of our approach lies in the proposed method of this computation. We show that by working backward in time, and using the discretized equations that are employed to compute the forward solution, the gradient computation parallels the forward computation in complexity and can be accomplished in $O(N)$ time. This is in contrast to the computationally intensive perturbation approach, which is widely used to compute the gradient, but requires N forward computations and is therefore $O(N^2)$. The method we propose is known as adjoint differentiation [100], which has been used to solve oceanographic and other computationally intensive inverse problems.

We present experimental results using simulated data to show that the proposed method results in superior quality reconstructions with substantial savings in computation.

5.2 Diffusion Data Model

The propagation of light through a medium can be described using Maxwell's equations. However, for the optical tomography problem, the electromagnetic formulation is not very useful due to the the intense scattering of photons in tissues. An alternative formulation, also capable of accounting for the effects of scattering, is transport theory. Transport theory treats propagating photons as particles, which is a valid simplification if the phases of the various contributions to the net scattered field are uncorrelated. This condition is satisfied in almost all cases of thick tissues illuminated by near infrared light. The governing equation describing the migration of photons that experience elastic scattering is the one-speed radiation transport equation [101, 102]. Under certain assumptions, the diffusion equation is a good approximation to the transport equation [103]. The approximation holds when the source is assumed to be isotropic and the medium has weak absorption. In this work, we assume that the measured data can be accurately modeled by the diffusion equation.

Let $U(x, y, t)$ be the intensity of light and $R(x, y, t)$ be the source strength at position (x, y) and time t. Let $\mu_a(x, y)$ and $\mu_s(x, y)$ denote the space-varying absorption and scattering coefficients. Let $D(x, y)$ denote the diffusion coefficient, given as

$$D(x, y) = \frac{c}{3[\mu_a(x, y) + (1 - g)\mu_s(x, y)]} \quad , \tag{5.1}$$

where c is the speed of the light in the medium and g is the scattering anisotropy parameter, equal to the average cosine of the scattering angle distribution[1]. Then the diffusion equation is given as

$$\frac{\partial U}{\partial t} = \frac{\partial}{\partial x} D \frac{\partial U}{\partial x} + \frac{\partial}{\partial y} D \frac{\partial U}{\partial y} - c\mu_a U + R \quad , \tag{5.2}$$

where the spatial and temporal dependence of the parameters has been suppressed. In this chapter, we parametrize the inverse problem in terms of D and μ_a. This is equivalent to recovering μ_a and μ_s due to Eq. (5.1).

5.2.1 Notation

Let S denote the set of discrete lattice points and let $s \in S$ denote the spatial position of a particular lattice point. In some instances we need to distinguish between the two spatial directions. In this case we use the subscript (i, j) to denote the spatial

[1] $\mu_s' = \mu_s(1 - g)$ is known as the effective scattering coefficient

position. When the spatial position subscript is present, the resulting quantity is a scalar with corresponding value at that spatial position (e.g. U_s or $U_{i,j}$). When the spatial position subscript is dropped, the resulting quantity is a column vector obtained by either row-ordering or column-ordering the corresponding two-dimensional field (e.g. U). We will use the superscript n to denote the discretized time index.

5.2.2 Measurement model

Let $\mathcal{M}$ denote the set of detector positions and $\mathcal{T}$ denote the set of time indices when the measurements are recorded. Let Y denote the measurements of the diffuse intensity U for all $s \in \mathcal{M}$ and $n \in \mathcal{T}$. In the interest of simplicity, we assume that the measurements are corrupted by uncorrelated Gaussian noise. However, the method we propose is not restricted to this choice. The log-likelihood of the observations Y given D and μ_a is

$$\log P(Y|D, \mu_a) = -\sum_{s \in \mathcal{M}} \frac{1}{2\sigma_s^2} \sum_{n \in \mathcal{T}} (Y_s^n - U_s^n)^2 \quad , \tag{5.3}$$

where σ_s^2 is the noise variance at spatial position s.

5.2.3 Computation of $\log P(Y|D, \mu_a)$

To compute the log-likelihood of the measurements Y given D and μ_a, we need to solve the diffusion equation (5.2) forward in time to obtain the diffuse intensity U_s^n for all time $n \in \mathcal{T}$ and spatial positions $s \in \mathcal{M}$. We propose to do this using a finite-difference approach where the spatial and temporal derivatives in Eq. (5.2) are replaced by their finite-difference approximations as follows

$$\frac{\partial}{\partial x} D \frac{\partial U}{\partial x} \approx \frac{D_{i+1/2,j}(U_{i+1,j} - U_{i,j}) - D_{i-1/2,j}(U_{i,j} - U_{i-1,j})}{\Delta^2} \triangleq \delta_x(U_{i,j}) \tag{5.4}$$

$$\frac{\partial}{\partial y} D \frac{\partial U}{\partial y} \approx \frac{D_{i,j+1/2}(U_{i,j+1} - U_{i,j}) - D_{i,j-1/2}(U_{i,j} - U_{i,j-1})}{\Delta^2} \triangleq \delta_y(U_{i,j}) \tag{5.5}$$

$$\frac{\partial U}{\partial t} \approx \frac{U^{n+1} - U^n}{\Delta t} \quad , \tag{5.6}$$

where $\Delta = \Delta x = \Delta y$ is assumed for simplicity. The approximations are obtained by simply differencing the second partial derivatives and centering each term appropriately. In doing so, we require the interpolated value of D halfway between the grid points. Linear interpolation is used to achieve this, i.e. $D_{i+1/2,j} = (D_{i,j} + D_{i+1,j})/2$.

By substituting the finite-difference approximations (5.4-5.6) in the diffusion equation (5.2), we obtain a difference equation that needs to be solved forward in time. When solving the difference equation for U^{n+1}, the finite-difference approximations to the spatial derivatives (5.4,5.5) can be evaluated at time index $n+1$ or n. The three methods we discuss in the following sections differ in this choice of the time index.

5.2.3.1 Explicit method

In this method, the spatial derivatives (5.4,5.5) are evaluated at the past time instance n when computing the diffuse intensity U^{n+1}. This is illustrated in Figure 5.1(a). Substituting Eqs. (5.4-5.6) in Eq. (5.2) and using time index n for the spatial derivatives, we obtain

$$U_{i,j}^{n+1} = (1 - c\mu_a \Delta t) U_{i,j}^{n} + \delta_x(U_{i,j}^{n})\Delta t + \delta_y(U_{i,j}^{n})\Delta t + (R_{i,j}^{n+1} + R_{i,j}^{n})\frac{\Delta t}{2} \quad . \tag{5.7}$$

Rewriting the above equation in vector-matrix notation, we obtain the following forward recursion

$$U^{n+1} = B\,U^{n} + \bar{R}^{n+1/2} \quad , \qquad \text{for } n = 0, 1, \ldots, T-1, \tag{5.8}$$

where U^{n+1} and U^{n} are row-ordered column vectors, B is a sparse matrix with four off-diagonal elements (corresponding to the vertical and horizontal neighbors of a pixel), T denotes the total time for which measurements are recorded, and $\bar{R}^{n+1/2}$ denotes the integrated source strength between time instances n and $n+1$. Using Eq. (5.8), we can compute U^{n} for any n by starting at $n = 0$ and moving forward in time. Furthermore, since B is a sparse matrix, the forward computation is $O(N)$, where N is the total number of discretized spatial positions. However, the disadvantage of this simple method is that it becomes unstable when [99]

$$\Delta t > \frac{\Delta^2}{4(\max_{s\in S} D_s)} \quad .$$

This method is not very useful because it can dictate very small time steps in the forward simulation.

5.2.3.2 Implicit method

In this method, the spatial derivatives (5.4,5.5) are evaluated at the present time instance $(n+1)$ when computing the diffuse intensity U^{n+1}. This is illustrated in Figure 5.1(b). Substituting Eqs. (5.4-5.6) in Eq. (5.2) and using time index $n+1$ for the spatial derivatives, we obtain

$$(1 + c\mu_a \Delta t)U_{i,j}^{n+1} - \delta_x(U_{i,j}^{n+1})\Delta t - \delta_y(U_{i,j}^{n+1})\Delta t = U_{i,j}^{n} + (R_{i,j}^{n+1} + R_{i,j}^{n})\frac{\Delta t}{2} \quad . \tag{5.9}$$

Rewriting the above equation in vector-matrix notation, we obtain the following forward recursion

$$A\, U^{n+1} = U^n + \bar{R}^{n+1/2} \quad , \qquad \text{for } n = 0, 1, \ldots, T-1, \tag{5.10}$$

where A is a sparse matrix having exactly the same structure as the B matrix in Sect. 5.2.3.1. The advantage of this method is that it is unconditionally stable for any value of Δt. However, the computation of U^{n+1} from U^n now requires the inversion of matrix A, which is not trivial anymore in terms of computation. Multigrid relaxation methods [104] can be employed to efficiently solve Eq. (5.10) but we prefer to use the method described in the next section.

5.2.3.3 Alternating-directions implicit method (ADI)

In this method, the computation of U^{n+1} from U^n is broken up in two time steps [99]. In the first half time step, only the spatial derivative in one direction is evaluated at the present time instance (implicit) and the other spatial derivative is evaluated at the previous time instance (explicit). In the next half time step, the implicit and explicit directions are switched. Figure 5.2 illustrates the switching of the implicit

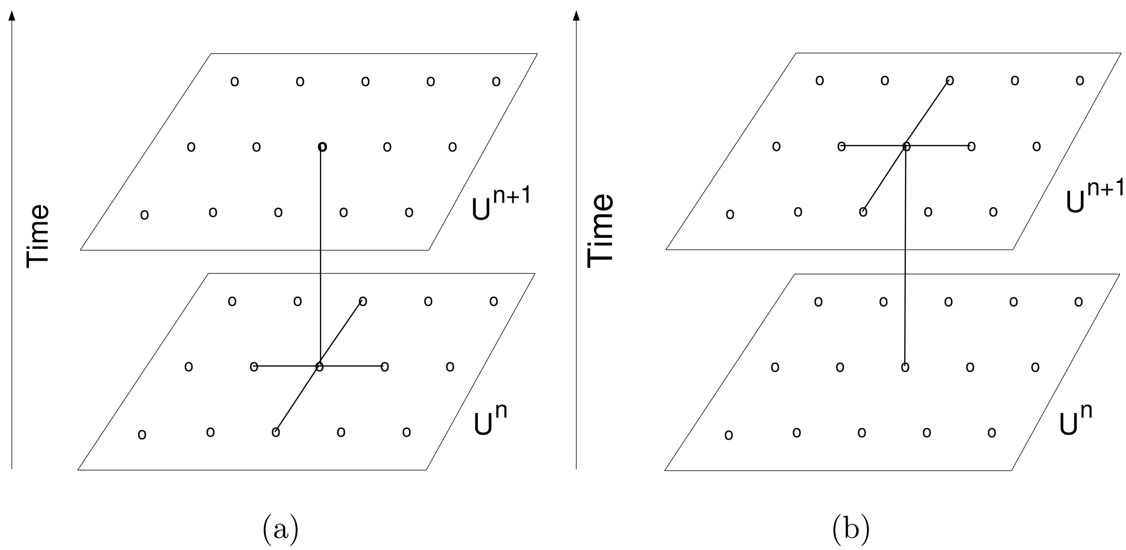


Fig. 5.1. Figure shows the time instance used to compute the finite-difference approximation of the spatial derivatives for the (a) explicit method and the (b) implicit method.

and explicit directions. The difference equations for the two half time steps are given as

$$(1 + c\mu_a \frac{\Delta t}{2})U_{i,j}^{n+1/2} - \delta_y(U_{i,j}^{n+1/2})\frac{\Delta t}{2} = U_{i,j}^{n} + \delta_x(U_{i,j}^{n})\frac{\Delta t}{2} + (R_{i,j}^{n+1} + R_{i,j}^{n})\frac{\Delta t}{4} \quad (5.11)$$

$$(1 + c\mu_a \frac{\Delta t}{2})U_{i,j}^{n+1} - \delta_x(U_{i,j}^{n+1})\frac{\Delta t}{2} = U_{i,j}^{n+1/2} + \delta_y(U_{i,j}^{n+1/2})\frac{\Delta t}{2} + (R_{i,j}^{n+1} + R_{i,j}^{n})\frac{\Delta t}{4} \ . \quad (5.12)$$

Let $\bar{R}^{n+1/4}$ denote the integrated source strength between time instances n and $n+1/2$. Then the difference Eqs. (5.11,5.12) for both halves of the time step can be compactly represented in vector-matrix notation by the single forward recursion

$$A\,U^{n+1/2} = B\,U^{n} + \bar{R}^{n+1/4} \ , \qquad \text{for } n = 0, 1/2, \ldots, T - 1/2, \qquad (5.13)$$

if we use row-ordering for U in Eq. (5.11), and column-ordering for U in Eq. (5.12). By switching the ordering as shown in Fig. 5.2, we force the structure of matrices A and B to remain the same while the absolute values differ for both halves of the time step. However, for the sake of notational simplicity, we will not distinguish between the two halves of each time step and use Eq. (5.13) to represent both of them.

To compute $U^{n+1/2}$ from U^n, we need to invert A. However, in this case A is always tridiagonal due the spatial derivative being implicit only in one direction. The

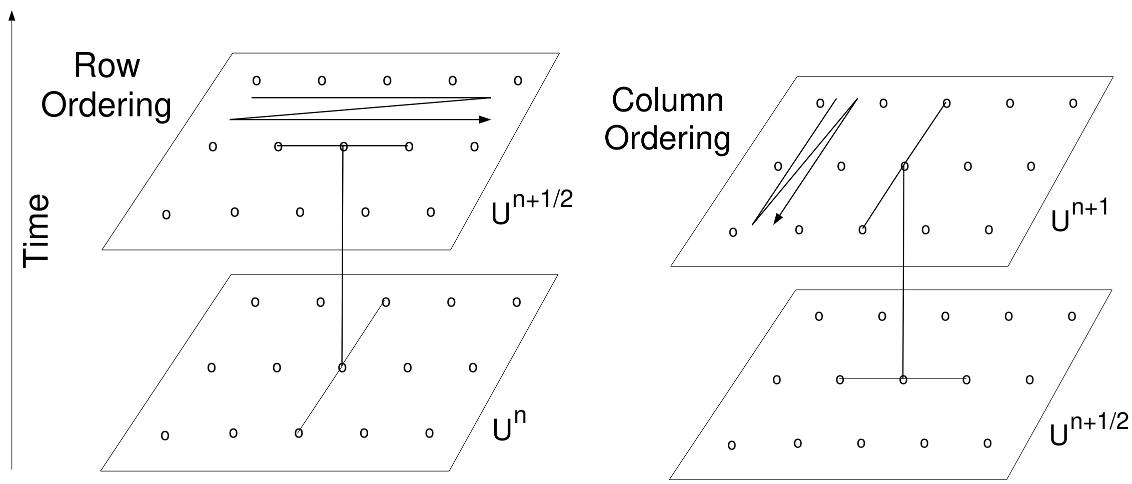


Fig. 5.2. Figure shows hows the implicit and explicit directions and the orderings are switched between the two halves of a single time step.

inversion can therefore be done in $O(N)$ computation. Furthermore, the method has been found to be stable in practice [98]. For the rest of the chapter, we assume that the ADI method will be used to do the forward computation.

5.2.4 Sensitivity computation

Let $\theta = [D\ \mu_a]^T$ be the column vector of the unknown parameters D and μ_a. Define

$$\phi(\theta) \triangleq \log P(Y|\theta) \quad .$$

To facilitate the solution of the inverse problem, we require the derivative or sensitivity of $\phi(\theta)$ with respect to θ. The technique that we propose for this computation is known as adjoint differentiation [100]. The method requires us to work backwards in time using the same discretized equations that were used to compute the forward solution in Sect. 5.2.3.3.

5.2.4.1 Sensitivity computation with respect to θ

The parameter θ is not a function of time n. However, to simplify the computation of the sensitivity of ϕ with respect to θ, we assume that θ varies with n. Let θ^n denote the parameter vector at time n. Then the sensitivity of ϕ with respect to θ_r^n is obtained by applying the chain rule

$$\frac{d\phi}{d\theta_r^n} = \sum_{s\in S} \frac{d\phi}{dU_s^{n+1/2}} \frac{dU_s^{n+1/2}}{d\theta_r^n} \qquad r = 1, \ldots, 2N.$$

Rewriting the above equation in vector-matrix notation, we obtain

$$\frac{d\phi}{d\theta^n} = \left[\frac{dU^{n+1/2}}{d\theta^n}\right]^T \frac{d\phi}{dU^{n+1/2}} \quad , \tag{5.14}$$

where

$$\frac{dU^{n+1/2}}{d\theta^n} = \left[\frac{dU^{n+1/2}}{d\theta_1^n}\ \frac{dU^{n+1/2}}{d\theta_2^n}\ \ldots\ \frac{dU^{n+1/2}}{d\theta_{2N}^n}\right] \quad . \tag{5.15}$$

Taking the derivative of Eq. (5.13) with respect to θ_r^n, we obtain

$$\frac{dA}{d\theta_r^n} U^{n+1/2} + A\frac{dU^{n+1/2}}{d\theta_r^n} = \frac{dB}{d\theta_r^n} U^n$$

$$\frac{dU^{n+1/2}}{d\theta_r^n} = A^{-1} \overbrace{\left(\frac{dB}{d\theta_r^n} U^n - \frac{dA}{d\theta_r^n} U^{n+1/2}\right)}^{X_r} \quad . \tag{5.16}$$

Using Eqs. (5.15) and (5.16), we obtain

$$\frac{dU^{n+1/2}}{d\theta} = A^{-1}[X_1\ X_2\ \ldots\ X_{2N}] = A^{-1}X \quad . \tag{5.17}$$

Substituting Eq. (5.17) in Eq. (5.14), we obtain

$$\frac{d\phi}{d\theta^n} = X^T (A^{-1})^T \frac{d\phi}{dU^{n+1/2}} \quad .$$

Finally, the total sensitivity of ϕ with respect to θ is obtained by setting $\theta^n = \theta$ for all n. Application of the chain rule then yields

$$\begin{aligned} \frac{d\phi}{d\theta} &= \sum_n \frac{d\phi}{d\theta^n} \frac{d\theta^n}{d\theta} \\ &= \sum_n X^T (A^{-1})^T \frac{d\phi}{dU^{n+1/2}} \quad . \end{aligned} \tag{5.18}$$

This computation can be done in $O(N)$ time since A is tridiagonal and X is a sparse matrix. The next section deals with the computation of $\frac{d\phi}{dU^n}$ for all n required in Eq. 5.18.

5.2.4.2 Sensitivity computation with respect to U

The sensitivity of ϕ with respect to U^n is obtained recursively by using the sensitivity of ϕ with respect to $U^{n+1/2}$. Application of the chain rule yields

$$\begin{aligned} \frac{d\phi}{dU_s^n} &= \sum_{r\in S} \frac{d\phi}{dU_r^{n+1/2}} \frac{dU_r^{n+1/2}}{dU_s^n} + \frac{\partial \phi}{\partial U_s^n} \\ &= \left[\frac{dU^{n+1/2}}{dU_s^n}\right]^T \frac{d\phi}{dU^{n+1/2}} + \frac{\partial \phi}{\partial U_s^n} \quad \text{for all } s \in S, \end{aligned} \tag{5.19}$$

where $\frac{\partial \phi}{\partial U_s^n}$ denotes the change in ϕ when only U_s^n is varied keeping all other variables constant. Differentiating Eq. (5.3) with respect to U_s^n, we obtain

$$\frac{\partial \phi}{\partial U_s^n} = \begin{cases} \frac{1}{\sigma_s^2}(Y_s^n - U_s^n) & s \in \mathcal{M},\ n \in \mathcal{T} \\ 0 & \text{otherwise} \end{cases} .$$

Rewriting Eq. (5.19) in vector-matrix notation, we obtain

$$\frac{d\phi}{dU^n} = \left[\frac{dU^{n+1/2}}{dU^n}\right]^T \frac{d\phi}{dU^{n+1/2}} + \frac{\partial \phi}{\partial U^n} \ , \tag{5.20}$$

where the matrix $\frac{dU^{n+1/2}}{dU^n}$ is given as

$$\frac{dU^{n+1/2}}{dU^n} = \left[\frac{dU^{n+1/2}}{dU_{s_1}^n} \ \frac{dU^{n+1/2}}{dU_{s_2}^n} \ \cdots \ \frac{dU^{n+1/2}}{dU_{s_N}^n}\right] \quad \{s_1, \ldots, s_N\} \in S. \tag{5.21}$$

Differentiating Eq. (5.13) with respect to U_s^n, we obtain

$$\frac{dU^{n+1/2}}{dU_s^n} = A^{-1} B\, I_s \quad \text{for all } s \in S \ , \tag{5.22}$$

where I_s is column vector that is zero everywhere except at spatial point s, where it is unity. Using Eqs. (5.22), (5.21) and (5.20), we obtain the sensitivity of ϕ with respect to U^n as

$$\frac{d\phi}{dU^n} = B^T (A^{-1})^T \frac{d\phi}{dU^{n+1/2}} + \frac{\partial \phi}{\partial U^n} \ . \tag{5.23}$$

Since A is tridiagonal and B is a sparse matrix with only two off-diagonal elements, the above computation can be done in $O(N)$ time.

By starting at the last time instance and using the recursion in Eq. (5.23), the contribution at each time step can be accumulated using Eq. (5.18) to obtain the complete sensitivity with respect to θ in $O(N)$ time.

5.3 Image Model

The problem of reconstructing the unknown parameters D and μ_a from the measurements Y is an ill-posed inverse problem and some form of regularization is necessary to make the solution well behaved. This is accomplished by incorporating an image model in the reconstruction process, which models our *a priori* knowledge regarding the unknown fields D and μ_a. The following development is for D; however, the development for μ_a follows the same pattern.

Markov random fields (MRF) have been extensively used in image processing applications since they yield a rich class of models with relatively simple parameterization. With this in view, we model D as a MRF or, equivalently, a Gibbs distribution with the following form

$$P(D) = \begin{cases} \frac{1}{z(\sigma_D)} \exp\{-u(D/\sigma_D)\} & \text{if } D \geq 0 \\ 0 & \text{otherwise} \end{cases} , \tag{5.24}$$

where σ_D is the scale parameter of the model and $z(\cdot)$ is the normalizing constant of the distribution known as the partition function. We impose a nonnegativity constraint on D because negative values of D are physically meaningless. We restrict the energy function $u(\cdot)$ to be of the form

$$u(\frac{D}{\sigma_D}) = \sum_{\{s,r\}\in\mathcal{N}} b_{s-r}\rho(\frac{D_s - D_r}{\sigma_D}) , \tag{5.25}$$

where $\mathcal{N}$ is the set of all neighboring pixel pairs, and $\rho(\cdot)$ is the potential function that assigns a cost to differences between neighboring pixel values.

A popular choice for $\rho(\cdot)$ in the signal-processing literature has been the quadratic function

$$\rho(\frac{\Delta}{\sigma}) = \frac{1}{2}\left|\frac{\Delta}{\sigma}\right|^2 .$$

This choice is similar to regularization methods that have been used earlier in optical tomography that impose constraints on the derivatives of the solution to enforce smoothness [94]. However, the quadratic cost function tends to excessively penalize large pixel differences resulting in blurred edges. This will be demonstrated in the results section.

Alternatively, a host of edge-preserving potential functions have been suggested in the literature [3, 10, 105, 12]. In particular, the generalized Gaussian MRF (GGMRF) [12] uses the following potential function

$$\rho(\Delta, \sigma) = \frac{1}{p}\left|\frac{\Delta}{\sigma}\right|^p \quad . \tag{5.26}$$

Note that when $p = 2$, the potential function is quadratic and the model reduces to a Gaussian MRF (GMRF). The advantage of using the GGMRF as the image model is two fold: first, it has been shown to provide good edge-preservation in the reconstructed image[12] for $p \approx 1$; second, the form of the model facilitates the estimation of the scale parameter directly from the data as shown in Chapter 2.

Using Eqs. (5.24), (5.25), and (5.26), the log-likelihood of the image D is given as

$$\log P(D) = -\frac{1}{p\sigma_D^p} \sum_{\{s,r\}\in\mathcal{N}} b_{s-r} \left|D_s - D_r\right|^p \quad . \tag{5.27}$$

The derivative of the log-likelihood with respect to D is given as

$$\frac{d \log P(D)}{dD_s} = \frac{1}{\sigma^p} \sum_{r\in\mathcal{N}_s} \text{sign}(D_r - D_s) b_{s-r} \left|D_s - D_r\right|^{p-1} \text{ for all } s \in S, \tag{5.28}$$

where $\mathcal{N}_s$ is the neighborhood of pixel s. The model for the absorption coefficient can be obtained by substituting σ_{μ_a} for σ_D and μ_a for D in Eqs. (5.27) and (5.28).

5.4 Reconstruction Criterion

We use the maximum *a posteriori* (MAP) estimation criterion to compute the reconstructions of D and μ_a. The MAP estimate is defined as

$$[\hat{D}, \hat{\mu}_a] = \arg \max_{[D,\mu_a]\geq 0} \log P(D, \mu_a|Y) \quad .$$

Using Bayes rule in the above equation, we have

$$[\hat{D}, \hat{\mu}_a] = \arg \max_{[D,\mu_a]\geq 0} \{\log P(Y|D, \mu_a) + \log P(D) + \log P(\mu_a)\} \quad .$$

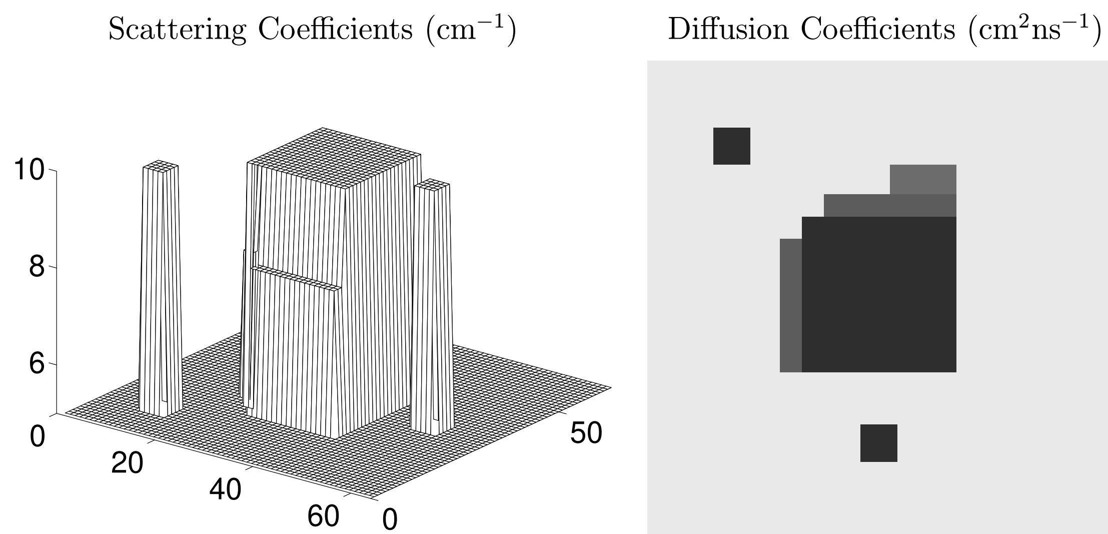
Since we can compute the derivative of each of the terms in the cost function (as shown in Sects. 5.2.4 and 5.3), we can use gradient-based methods to do the above optimization. In particular, we use the conjugate gradient algorithm because of its superior convergence properties. However, conventional CGD cannot be used directly since it requires line searching that can violate the positivity constraint. We modify the search procedure so that the estimate is projected back on the convex set $[D\ \mu_a] \geq 0$ when searching for a minimum in a particular direction. This procedure is referred to as bent-line searching.

5.5 Results

In this section, we study the performance of the proposed algorithm using simulated data. Although the method developed in this chapter can be used to estimate D and μ_a simultaneously, we will restrict ourselves to the simple case of just estimating D and assume μ_a is known.

We will use a GGMRF with $p = 1.1$ as the image model since it results in good edge-preservation. We will also show the reconstruction corresponding to $p = 2.0$ (GMRF) to compare the quality of our reconstruction to previously used constraint-based regularization methods that use a quadratic penalty[94]. We use an 8 point neighborhood for the MRF with $b_{s-r} = (2\sqrt{2}+4)^{-1}$ for nearest neighbors and $b_{s-r} = (4\sqrt{2}+4)^{-1}$ for diagonal neighbors. Ideally, the scale parameter σ_D needs to be estimated directly from the measurements Y before computing the unknown field D. This unsupervised estimation of σ_D is a challenging problem in itself and is not within the scope of this research. Therefore, for the purpose of this research, we will fix the value of σ_D to the ML estimate (refer to Chapter 2) obtained from the original D, which is known in our simulation examples.

The values of the absorption and scattering coefficients used in our examples have been chosen to reflect those of real tissues. The physical dimensions have also been chosen to reflect a real medical imaging set-up. Figure 5.3(a) shows the scattering coefficients chosen for the simulation example on a 64x64 grid with 0.1 cm spatial separation between grid points. The absorption coefficients were set to 0.1 cm^{-1} over the entire field. Figure 5.3(b) shows the diffusion coefficients computed from the absorption and scattering coefficients using $c = 22$ cm/ns as the speed of light in the medium. Figures 5.4(a-c) show the different source detector configurations that we use for this example. The white dots mark the source positions and the black dots mark the detector locations. The data are collected by turning on one source at a time and making measurements at all the detector locations for all $n \in \mathcal{T}$. The data collected for each active source constitutes a single view of the object. Let the data collected in view v be denoted as $Y^{(v)}$. Let $\phi^{(v)} \triangleq \log P(Y^{(v)}|\theta)$. Then the total log-likelihood of the data for multiple views is given as $\phi \triangleq \sum_v \phi^{(v)}$, where each $\phi^{(v)}$ is computed using the procedure described in Sect. 5.2.3.3. Similarly, the total sensitivity of ϕ with respect to θ is given as $\frac{d\phi}{d\theta} = \sum_v \frac{d\phi^{(v)}}{d\theta}$, where each $\frac{d\phi^{(v)}}{d\theta}$ is computed using the procedure in Sect. 5.2.4.



(a) (b)

Fig. 5.3. Figure shows the (a) scattering coefficients (μ_s) and (b) diffusion coefficients (D) chosen for the simulation example. The absorption coefficients (μ_a) was set to 0.1 cm^{-1} over the entire field.

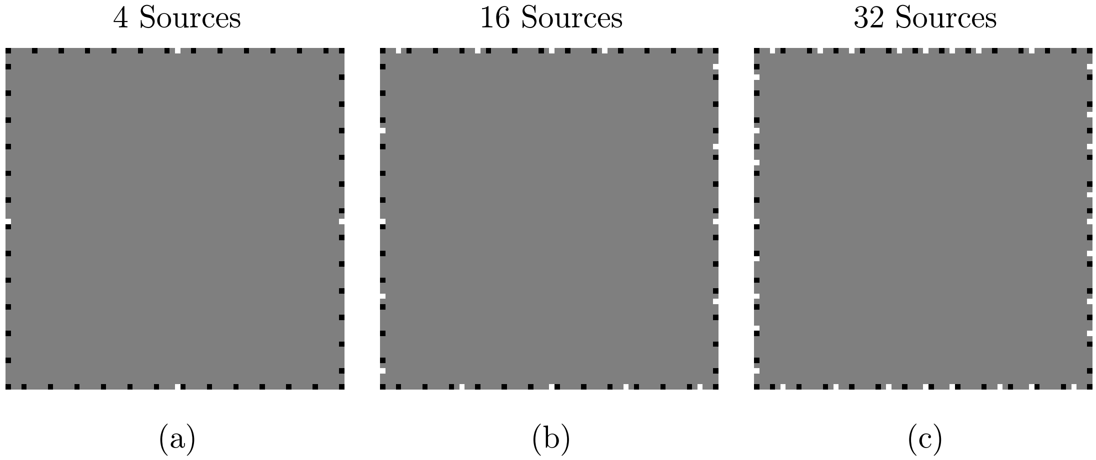
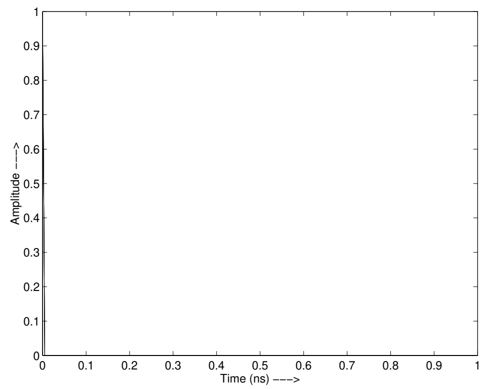


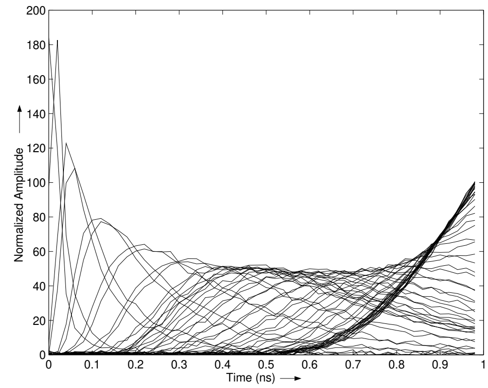
Fig. 5.4. Configurations used for the sources and detectors. 52 detectors are placed on the periphery of the object with (a) 4 sources, (b) 16 sources, and (c) 32 sources. Note that the sources in (a) are contained in (b), and the sources in (b) are contained in (c).

Figure 5.5(a) shows the source distribution that we use for all the sources, which is a single pulse. Figure 5.5(b) shows all the detector responses for a single view when the SNR of the detectors is set at 30 dB. The simulation is done using $\Delta t = 0.005$ ns for a total time of $T = 1$ ns while the detector resolution is kept at 0.02 ns.

Figure 5.6 shows the MAP reconstructions of the diffusion coefficients using the conjugate gradient algorithm when 16 sources are used. The algorithm was initialized with a constant diffusion coefficient field with value 1.0 $\mathrm{cm}^2\mathrm{ns}^{-1}$. Figures 5.6(b) and 5.6(c) show the reconstructions corresponding to a GGMRF ($p = 1.1$) and a GMRF ($p = 2.0$) respectively. Notice that the GGMRF reconstruction has sharp edges and is a good reproduction of the original image considering the limited number of views. On the other hand, the GMRF reconstruction has blurred edges and excessive residual noise. Figure 5.7 shows the effect of using additional sources (views) in the MAP reconstructions. A GGMRF prior with $p = 1.1$ is used and the detector SNR is set at 30 dB. We see from Figs. 5.7(b) and 5.7(c) that there is considerable improvement going from 4 sources to 16 sources. However, increasing the number of sources to 32



(a)



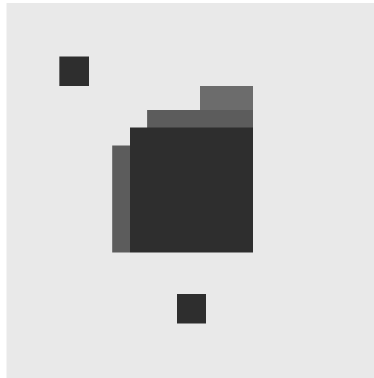
(b)

Fig. 5.5. (a) Source distribution used is a single pulse. (b) Normalized measurements obtained at all the detector positions with 30 dB SNR when one of the sources is active. The normalizations was done by the standard deviation of the noise present in each measurement.

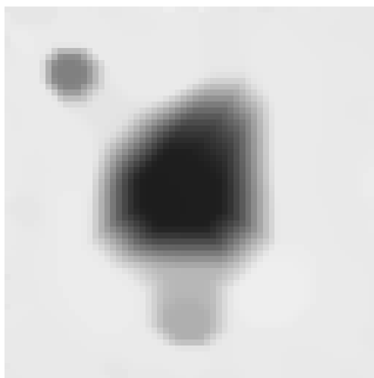
does not yield any additional improvement as seen in Fig. 5.7(d). Figure 5.8 shows the effect of noise in the detector responses in the MAP reconstructions. A GGMRF prior with $p = 1.1$ is used and 4 sources are used. We see from Fig. 5.8(d) that when the detector responses have 10 dB SNR, the two scatterers in the original image can no longer be resolved.

5.6 Conclusion

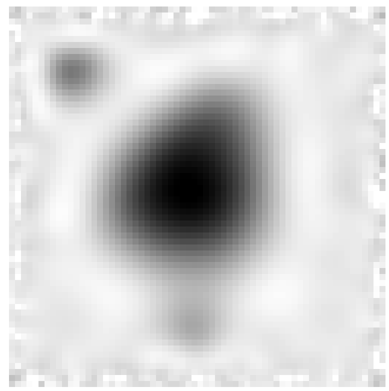
The contribution of the research in this chapter has been threefold. First, the use of a new alternating-directions implicit method to solve the forward diffusion problem in O(N) time. Second, the use of adjoint differentiation to compute the sensitivity of the measurements with respect to the unknown parameters. By working backwards in time, and using the discretized equations that are employed to compute the forward solution, we have shown that this can be done in $O(N)$ time as opposed to $O(N^2)$ time required by the perturbation approach. Third, an edge-preserving GGMRF model



(a)



(b)

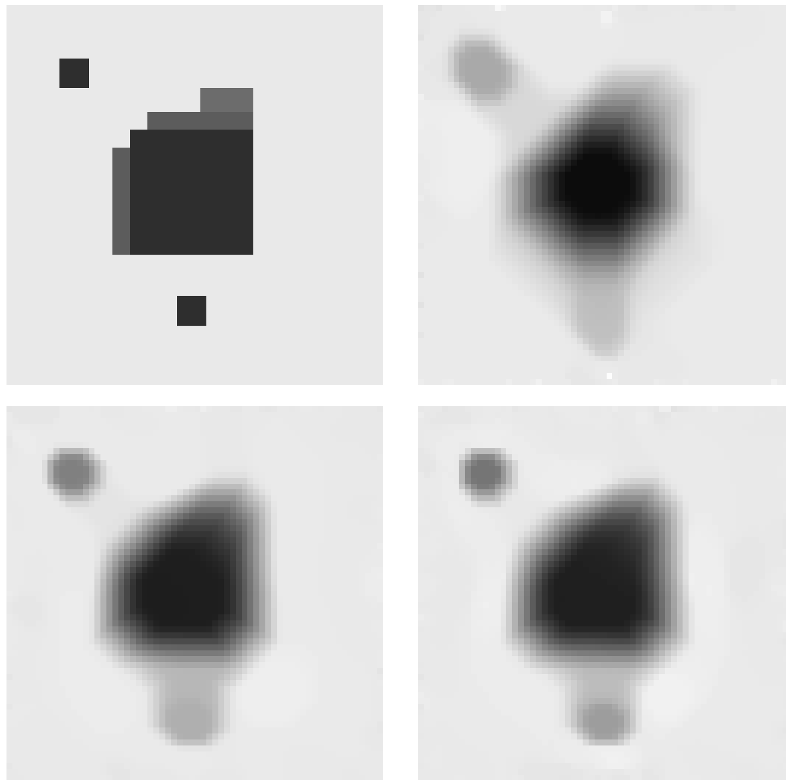


(c)

Fig. 5.6. (a) Original diffusion coefficients. Reconstructed diffusion coefficients when 16 sources are used for the GGMRF image model with (b) $p = 1.1$, and (c) $p = 2.0$. Note that white in the images corresponds to 1.52 $\mathrm{cm}^2\mathrm{ns}^{-1}$ and black corresponds to 0.55 $\mathrm{cm}^2\mathrm{ns}^{-1}$.

has been employed and shown to be superior to the standard quadratic regularization method, which results in blurred edges and excess residual noise.

Moreover, the proposed method of solving the inverse problem is very general and can be applied to complex problems such as geophysical structure estimation from seismic data and ocean surface reconstruction using sonar. The only requirement for the applicability of the proposed method is that the measurements for the process in question be adequately described by a predictive forward computational model.

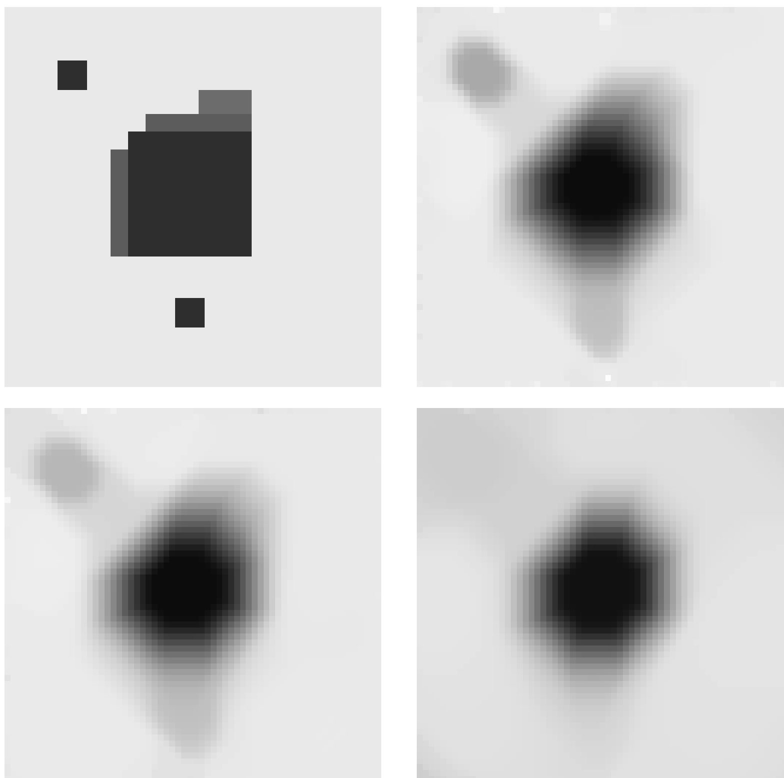


(a) (b) (c) (d)

Fig. 5.7. (a) Original diffusion coefficients. Reconstructed diffusion coefficients when the prior model is a GGMRF $p = 1.1$ and the number of sources are (b) 4, (c) 16, and (d) 32. Note that white in the images corresponds to 1.52 $\mathrm{cm}^2\mathrm{ns}^{-1}$ and black corresponds to 0.55 $\mathrm{cm}^2\mathrm{ns}^{-1}$.

It can also be used to optimize engineering designs in complex situations such as streamlining of airplane foils and automobile bodies to reduce drag.

Future research should focus on multiresolution strategies for computing the MAP estimate, employing multiscale image models proposed in Chapter 4 as opposed to fixed-resolution image models to describe the unknown image, and unsupervised methods for optimal estimation of image and data model parameters.



(a) (b) (c) (d)

Fig. 5.8. (a) Original diffusion coefficients. Reconstructed diffusion coefficients when the prior model is a GGMRF $p = 1.1$, the number of sources are 4, and the detector SNR is (b) 30dB, (c) 20dB, and (d) 10dB. Note that white in the images corresponds to 1.52 cm^2ns^{-1} and black corresponds to 0.55 cm^2ns^{-1}.

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] S. Geman and D. McClure, "Bayesian images analysis: An application to single photon emission tomography," *Proc. Statist. Comput. sect. Amer. Stat. Assoc.*, 1985, Washington, DC, pp. 12–18.

[2] T. Hebert and R. Leahy, "A generalized EM algorithm for 3-d Bayesian reconstruction from Poisson data using Gibbs priors," *IEEE Trans. on Medical Imaging*, vol. 8, no. 2, pp. 194–202, June 1989.

[3] P. J. Green, "Bayesian reconstruction from emission tomography data using a modified EM algorithm," *IEEE Trans. on Medical Imaging*, vol. 9, no. 1, pp. 84–93, March 1990.

[4] K. Lange and R. Carson, "EM reconstruction algorithms for emission and transmission tomography," *J. Comput. Assist. Tomography*, vol. 8, no. 2, pp. 306–316, April 1984.

[5] J. M. Ollinger, "Maximum-likelihood reconstruction of transmission images in emission computed tomography via the EM algorithm," *IEEE Trans. on Medical Imaging*, vol. 13, no. 1, pp. 89–101, March 1994.

[6] K. Sauer and C. A. Bouman, "A local update strategy for iterative reconstruction from projections," *IEEE Trans. on Signal Processing*, vol. 41, no. 2, pp. 534–548, February 1993.

[7] B. Hunt, "Bayesian methods in nonlinear digital image restoration," *IEEE Trans. on Comput.*, vol. c-26, no. 3, pp. 219–229, 1977.

[8] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. PAMI-6, pp. 721–741, Nov. 1984.

[9] A. Blake and A. Zisserman, *Visual Reconstruction.* Cambridge, Massachusetts: MIT Press, 1987.

[10] R. Stevenson and E. Delp, "Fitting curves with discontinuities," *Proc. of the first international workshop on robust computer vision*, pp. 127–136, October 1-3 1990.

[11] D. Geman and G. Reynolds, "Constrained restoration and the recovery of discontinuities," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 14, no. 3, pp. 367–383, 1992.

[12] C. A. Bouman and K. Sauer, "A generalized Gaussian image model for edge-preserving map estimation," *IEEE Trans. on Image Processing*, vol. 2, pp. 296–310, July 1993.

[13] S. Brette, J. Idier, and A. Mohammad-Djafari, "Scale invariant Markov models for Bayesian inversion of linear inverse problems," in *Maximum Entropy and Bayesian Methods* (J. Skilling and S. Sibisi, eds.), pp. 199–212, Netherlands: Kluwer Academic Publishers, 1996.

[14] L. E. Baum and T. Petrie, "Statistical inference for probabilistic functions of finite state Markov chains," *Ann. Math. Statistics*, vol. 37, pp. 1554–1563, 1966.

[15] L. Baum, T. Petrie, G. Soules, and N. Weiss, "A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains," *Ann. Math. Statistics*, vol. 41, no. 1, pp. 164–171, 1970.

[16] L. Shepp and Y. Vardi, "Maximum likelihood reconstruction for emission tomography," *IEEE Trans. on Medical Imaging*, vol. MI-1, no. 2, pp. 113–122, October 1982.

[17] D. Pickard, "Asymptotic inference for an ising lattice iii. non-zero field and ferromagnetic states," *J. Appl. Prob.*, vol. 16, pp. 12–24, 1979.

[18] D. Pickard, "Inference for discrete Markov fields: The simplest nontrivial case," *Journal of the American Statistical Association*, vol. 82, pp. 90–96, March 1987.

[19] J. Besag, "On the statistical analysis of dirty pictures," *Journal of the Royal Statistical Society B*, vol. 48, no. 3, pp. 259–302, 1986.

[20] S. Geman and C. Graffigne, "Markov random field image models and their applications to computer vision," *Proc. of the Intl Congress of Mathematicians*, 1986, Berkeley, California, pp. 1496–1517.

[21] S. Lakshmanan and H. Derin, "Simultaneous parameter estimation and segmentation of Gibbs random fields using simulated annealing," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 11, no. 8, pp. 799–813, August 1989.

[22] B. Chalmond, "An iterative Gibbsian technique for reconstruction of m-ary images," *Pattern Recognition*, vol. 22, pp. 747–761, 1989.

[23] Y. Ogata, "A Monte Carlo method for an objective Bayesian procedure," *Ann. Inst. of Statist. Math.*, vol. 42, no. 3, pp. 403–433, 1990.

[24] C. Geyer and E. Thompson, "Constrained monte carlo maximum likelihood for dependent data," *Journal of the Royal Statistical Society B*, vol. 54, pp. 657–699, 1992.

[25] B. Gidas, "Parameter estimation for Gibbs distributions from fully observed data," in *Markov Random Fields: Theory and Applications* (R. Chellappa and A. Jain, eds.), pp. 471–499, Boston: Academic Press, Inc., 1993.

[26] C. Geyer, "On the convergence of Monte Carlo maximum likelihood calculations," *Journal of the Royal Statistical Society B*, vol. 56, pp. 261–274, 1994.

[27] G. G. Potamianos and J. K. Goutsias, "Partition function estimation of Gibbs random field images using Monte Carlo simmulation," *IEEE Trans. on Information Theory*, vol. 39, no. 4, pp. 1322–1332, July 1993.

[28] M. Jerrum and A. Sinclair, "Polynomial-time approximation algorithms for the ising model," *SIAM Journal of Computing*, vol. 22, no. 5, pp. 1087–1116, 1993.

[29] N. P. Galatsanos and A. K. Katsaggelos, "Methods for choosing the regularization parameter and estimating the noise variance in image restoration and their relation," *IEEE Trans. on Image Processing*, vol. 1, no. 3, pp. 322–336, July 1992.

[30] A. Mohammad-Djafari, "On the estimation of hyperparameters in Bayesian approach of solving inverse problems," *Proc. of IEEE Int'l Conf. on Acoust., Speech and Sig. Proc.*, April 27-30 1993, Minneapolis, Minnesota, pp. 495–498.

[31] K. Lange, "An overview of Bayesian methods in image reconstruction," *Proc. of the SPIE Conference on Digital Image Synthesis and Inverse Optics*, vol. SPIE-1351, 1990, San Diego, CA, pp. 270–287.

[32] Z. Zhou and R. Leahy, "Approximate maximum likelihood hyperparameter estimation for Gibbs priors," Tech. Rep. USE-SIPI 285, Signal and image processing institute, University of southern California, Los Angeles, CA 90089., June 1995.

[33] A. Mohammad-Djafari, "Joint estimation of parameters and hyperparameters in a Bayesian approach of solving inverse problems," *Proc. of IEEE Int'l Conf. on Image Proc.*, vol. II, September 16-19 1996, Lausanne, Switzerland, pp. 473–476.

[34] R. Schultz, R. Stevenson, and A. Lumsdaine, "Maximum likelihood parameter estimation for non-Gaussian prior signal models," *Proc. of IEEE Int'l Conf. on Image Proc.*, vol. 2, November 1994, Austin, TX, pp. 700–704.

[35] D. M. Higdon, V. E. Johnson, J. E. Bowsher, T. G. Turkington, D. R. Gilland, and R. J. Jaszczack, "Fully Bayesian estimation of Gibbs hyperparameters for emission computed tomography data," Technical Report 96-21, Institute of Statistics and Decision Sciences, Duke University, 1995.

[36] Z. Zhou and R. Leahy, "Approximate maximum likelihood hyperparameter estimation for Gibbs priors," *Proc. of IEEE Int'l Conf. on Image Proc.*, October 23-26 1995, Washington, D.C., pp. 284–287.

[37] J. Zhang, "The mean field theory in EM procedures for Markov random fields," *IEEE Trans. on Signal Processing*, vol. 40, no. 10, pp. 2570–2583, October 1992.

[38] W. Pun and B. Jeffs, "Shape parameter estimation for generalized Gaussian Markov random field models used in MAP image restoration," *29th Asilomar Conference on Signals, Systems, and Computers*, Oct. 29 - Nov. 1 1995.

[39] B. D. Jeffs and W. H. Pun, "Simple shape parameter estimation from blurred observations for a generalized Gaussian MRF image prior used in MAP image restoration," *Proc. of IEEE Int'l Conf. on Image Proc.*, September 16-19 1996, Lausanne, Switzerland, pp. 465–468.

[40] S. Geman and D. McClure, "Statistical methods for tomographic image reconstruction," *Bull. Int. Stat. Inst.*, vol. LII-4, pp. 5–21, 1987.

[41] C. A. Bouman and K. Sauer, "Maximum likelihood scale estimation for a class of Markov random fields," *Proc. of IEEE Int'l Conf. on Acoust., Speech and Sig. Proc.*, vol. 5, April 19-22 1994, Adelaide, South Australia, pp. 537–540.

[42] C. A. Bouman, K. Sauer, and S. S. Saquib, "Tractable models and efficient algorithms for Bayesian tomography," *Proc. of IEEE Int'l Conf. on Acoust., Speech and Sig. Proc.*, May 9-12 1995, Detroit, MI, pp. 2907–2910.

[43] S. S. Saquib, C. A. Bouman, and K. Sauer, "Efficient ML estimation of the shape parameter for generalized Gaussian MRF," *Proc. of IEEE Int'l Conf. on Acoust., Speech and Sig. Proc.*, vol. 4, May 7-10 1996, Atlanta, GA, pp. 2229–2232.

[44] S. S. Saquib, C. A. Bouman, and K. Sauer, "ML parameter estimation for Markov random fields, with applications to Bayesian tomography," Tech. Rep. TR-ECE 95-24, School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907, October 1995.

[45] K. Sauer and C. A. Bouman, "Maximum likelihood dosage estimation for Bayesian tomography," *Proc. of IEEE Int'l Conf. on Image Proc.*, vol. 2, Nov. 13-16 1994, Austin, Texas, pp. 844–848.

[46] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller, "Equations of state calculations by fast computing machines," *J. Chem. Phys.*, vol. 21, pp. 1087–1091, 1953.

[47] W. K. Hastings, "Monte Carlo sampling methods using Markov chains and their applications," *Biometrika*, vol. 57, no. 1, pp. 97–109, 1970.

[48] P. H. Peskun, “Optimum Monte-Carlo sampling using Markov chains,” *Biometrika*, vol. 60, no. 3, pp. 607–612, 1973.

[49] P. J. Green and X. liang Han, “Metropolis methods, Gaussian proposals and antithetic variables,” in *Stochastic Models, Statistical methods, and Algorithms in Image Analysis* (P. Barone, A. Frigessi, and M. Piccioni, eds.), pp. 142–164, Berlin: Springer-Verlag, 1992.

[50] C. A. Bouman and K. Sauer, “A unified approach to statistical tomography using coordinate descent optimization,” *IEEE Trans. on Image Processing*, vol. 5, no. 3, pp. 480–492, March 1996.

[51] A. Blake, “Comparison of the efficiency of deterministic and stochastic algorithms for visual reconstruction,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 11, no. 1, pp. 2–30, January 1989.

[52] J. Besag, “Towards Bayesian image analysis,” *Journal of Applied Statistics*, vol. 16, no. 3, pp. 395–407, 1989.

[53] J. A. O’Sullivan, “Divergence penalty for image regularization,” *Proc. of IEEE Int’l Conf. on Acoust., Speech and Sig. Proc.*, vol. 5, April 19-22 1994, Adelaide, South Australia, pp. 541–544.

[54] S. Ihara, *Information Theory for continuous systems.* Singapore: World Scientific, 1993.

[55] Y. Ogata and M. Tanemura, “Likelihood analysis of spatial point patterns,” *Journal of the Royal Statistical Society B*, vol. 46, no. 3, pp. 496–518, 1984.

[56] A. D. Pierro, “A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography,” *IEEE Trans. on Medical Imaging*, vol. 14, no. 1, pp. 132–137, 1994.

[57] K. D. Sauer, S. Borman, and C. A. Bouman, “Parallel computation of sequential pixel updates in statistical tomographic reconstruction,” *Proc. of IEEE Int’l Conf. on Image Proc.*, vol. 2, Oct. 22-25 1995, Washington, DC, pp. 93–96.

[58] J. Fessler, E. Ficaro, N. Clinthorne, and K. Lange, “Fast parallelizable algorithm for transmission image reconstruction,” *Proc. of IEEE Nucl. Sci. Symp. and Med. Imaging Conf.*, October 21-28 1995, San Francisco, California.

[59] J. Fessler and A. Hero, “Penalized maximum-likelihood image reconstruction using space-alternating generalized EM algorithms,” *IEEE Trans. on Image Processing*, vol. 4, no. 10, pp. 1417–1429, October 1995.

[60] E. Levitan and G. Herman, “A maximum a *posteriori* probability expectation maximization algorithm for image reconstruction in emission tomography,” *IEEE Trans. on Medical Imaging*, vol. MI-6, pp. 185–192, Sept. 1987.

[61] R. F. D. Paiva, *Developpement d'un Microtomographe X et Application a la caracterisation des roches reservoirs.* Ph.d. dissertation, Universite Paris VI, Paris, France, October 1995.

[62] D. Geiger and F. Girosi, "Parallel and deterministic algorithms from MRF's: Surface reconstruction," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. PAMI-13, no. 5, pp. 401–412, May 1991.

[63] Y.-L. You, M. Kaveh, W. Xu, and A. Tannenbaum, "Analysis and design of anisotropic diffusion for image processing," *Proc. of IEEE Int'l Conf. on Image Proc.*, vol. I, November 13-16 1994, Austin, TX, pp. 497–501.

[64] F. Catte, P.-L. Lions, J.-M. Morel, and T. Coll, "Image selective smoothing and edge detection by nonlinear diffusion," *SIAM Journal on Numerical Analysis*, vol. 29, no. 1, pp. 182–193, February 1992.

[65] P. Charbonnier, L. Blanc-Feraud, G. Aubert, and M. Barlaud, "Two deterministic half-quadratic reqularization algorithms for computed imaging," *Proc. of IEEE Int'l Conf. on Image Proc.*, vol. 2, November, 13-16 1994, Austin, TX, pp. 168–176.

[66] D. Geman and C. Yang, "Nonlinear image recovery with half-quadratic regularization," *IEEE Trans. on Image Processing*, vol. 4, no. 7, pp. 932–946, July 1995.

[67] L. Blanc-Feraud, S. Teboul, G. Aubert, and M. Barlaud, "Nonlinear regularization using constrained edges in image reconstrution," *Proc. of IEEE Int'l Conf. on Image Proc.*, vol. II, September, 16-19 1996, Lausanne, Switzerland, pp. 449–452.

[68] M. Nitzberg and T. Shiota, "Nonlinear image filtering with edge and corner enhancement," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 14, no. 8, pp. 826–833, August 1992.

[69] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 12, no. 7, pp. 629–639, July 1990.

[70] Y.-L. You, W. Xu, A. Tannenbaum, and M. Kaveh, "Behavioral analysis of anisotropic diffusion in image processing," *IEEE Trans. on Image Processing*, vol. 5, no. 11, pp. 1539–1553, November 1996.

[71] C. Chen, X. Ouyang, W. Wong, X. Hu, V. Johnson, C. Ordonez, and C. Metz, "Sensor fusion in image reconstruction," *IEEE Trans. on Nuclear Science*, vol. 38, pp. 687–692, 1991.

[72] P.-C. Chiao, W. Rogers, J. Fessler, N. Clinthorne, and A. Hero, "Model-based estimation with boundary side information or boundary regularization," *IEEE Trans. on Medical Imaging*, vol. 13, pp. 227–234, June 1994.

[73] S. S. Saquib, C. A. Bouman, , and K. Sauer, "A non-homogeneous MRF model for multiresolution Bayesian estimation," *Proc. of IEEE Int'l Conf. on Image Proc.*, vol. 2, September 16-19 1996, Lausanne Switzerland, pp. 445–448.

[74] S. S. Saquib, C. A. Bouman, and K. Sauer, "A new multiscale image model for Bayesian tomography," *Proc. of the Image and Multidimensional Signal Proc. Workshop*, March 3-6 1996, Belize City, Belize, pp. 144–145.

[75] C. A. Bouman and M. Shapiro, "A multiscale random field model for Bayesian image segmentation," *IEEE Trans. on Image Processing*, vol. 3, no. 2, pp. 162–177, March 1994.

[76] C. A. Bouman and B. Liu, "Multiple resolution segmentation of textured images," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 13, pp. 99–113, Feb. 1991.

[77] M. R. Luettgen, W. C. Karl, and A. S. Willsky, "Efficient multiscale regularization with applications to the computation of optical flow," *IEEE Trans. on Image Processing*, vol. 3, no. 1, pp. 41–64, January 1994.

[78] F. Heitz, P. Perez, and P. Bouthemy, "Multiscale minimization of global energy functions in some visual recovery problems," *Comput. Vision Graphics and Image Process.*, vol. 59, no. 1, pp. 125–134, January 1994.

[79] B. Sahiner and A. Yagle, "Image reconstruction from projections under wavelet constraints," *IEEE Trans. on Signal Processing*, vol. 41, no. 12, pp. 3579–3584, 1993.

[80] A. Delaney and Y. Bresler, "Multiresolution tomographic reconstruction using wavelets," *IEEE Trans. on Image Processing*, vol. 4, no. 6, pp. 799–813, June 1995.

[81] M. Bhatia, W. C. Karl, and A. S. Willsky, "Wavelet-based method for multiscale tomographic reconstruction," *IEEE Trans. on Medical Imaging*, vol. 15, no. 1, pp. 92–101, 1996.

[82] E. *Ü*. Mumcuoğlu, R. Leahy, S. R. Cherry, and Z. Zhou, "Fast gradient-based methods for Bayesian reconstruction of transmission and emission pet images," *IEEE Trans. on Medical Imaging*, vol. 13, no. 4, pp. 687–701, December 1994.

[83] R. G. Keys, "Cubic convolution interpolation for digital image processing," *IEEE Trans. on Acoust. Speech and Signal Proc.*, vol. 29, no. 6, pp. 1153–1160, December 1981.

[84] B. Horn and B. Schunck, "Determining optical flow," *Artificial Intelligence*, vol. 17, pp. 185–203, 1981.

[85] J. Konrad and E. Dubois, "Bayesian estimation of motion vector fields," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 14, no. 9, pp. 910–927, September 1992.

[86] B. Rouchouze, P. Mathieu, T. Gaidon, and M. Barlaud, "Motion estimation based on Markov random fields," *Proc. of IEEE Int'l Conf. on Image Proc.*, vol. 3, November, 13-16 1994, Austin, TX, pp. 270–274.

[87] J. Zhang and G. G. Hanauer, "The application of mean field theory to image motion estimation," *IEEE Trans. on Image Processing*, vol. 4, no. 1, pp. 19–33, January 1995.

[88] G. Muller, B. Chance, R. Alfano, S. Arridge, J. Beuthan, E. Gratton, M. Kaschke, B. Masters, S. Svanberg, and P. van der Zee, eds., *Medical Optical Tomography: Functional Imaging and Monitoring*, vol. IS 11. Bellingham, WA: SPIE press, 1993.

[89] S. Flock, M. Patterson, B. Wilson, and D. Wyman, "Monte Carlo modeling of light propagation in highly scattering tissues - I: Model predictions and comparison with diffusion theory," *IEEE Trans. Biomed. Eng.*, vol. 36, no. 12, pp. 1162–1168, 1989.

[90] Y. Wang, J. Chang, R. Aronson, R. Barbour, H. Graber, and J. Lubowsky, "Imaging of scattering media by diffusion tomography: an iterative perturbative approach," *Physiological Monitoring and Early Detection Diagnostic Methods*, vol. 1641, 1992, pp. 58–71.

[91] R. Barbour, H. Graber, Y. Wang, J. Chang, and R. Aronson, "A perturbation approach for optical diffusion tomography using continuous-wave and time-resolved data," in *Medical Optical Tomography: Functional Imaging and Monitoring*, vol. IS 11, pp. 87–120, SPIE Press, 1993.

[92] J. Chang, H. Graber, and R. Barbour, "Image reconstruction of targets in random media from continuous wave laser measurements and simulated data," *Advances in Optical Imaging and Photon Migration* (R. R. Alfano, ed.), vol. 21, 1994, pp. 193–201.

[93] B. B. Das, J. Dolne, R. L. Barbour, H. L. Graber, J. Chang, M. Zevallos, F. Liu, and R. R. Alfano, "Analysis of time-resolved data for tomographical image reconstruction of opaque phantoms and finite absorbers in diffusive media," vol. 2389, 1995, pp. 16–28.

[94] S. R. Arridge, "Forward and inverse problems in time-resolved infrared imaging," in *Medical Optical Tomography: Functional Imaging and Monitoring*, vol. IS 11, pp. 35–64, SPIE Press, 1993.

[95] R. Barbour, H. Graber, R. Aronson, and J. Lubowsky, "Imaging of subsurface regions of random media by remote sensing," vol. 1431, 1991, pp. 192–203.

[96] R. Model, R. Hunlich, M. Orlt, and M. Walzel, "Image reconstruction for random media by diffusion tomography," vol. 2389, 1995, pp. 400–410.

[97] D. S. Burnett, *Finite Element Analysis: From Concepts to Applications.* New York: Addison-Wesley, 1987.

[98] W. F. Ames, *Numerical Methods for Partial Differential Equations.* New York: Academic Press, 1977.

[99] W. Press, S. Teukolsky, W. Vetterling, and B. Flannery, *Numerical Recipes in C: The Art of Scientific Computing.* Cambridge: Cambridge University Press, second ed., 1992.

[100] K. M. Hanson and G. S. Cunningham, "A computational approach to Bayesian inference," in *Computing Science and Statistics* (M. M. Meyer and J. L. Rosenberger, eds.), pp. 202–211, Fairfax Station, VA 22039-7460: Interface Foundation, 1996.

[101] M. C. Case and P. F. Zweifel, *Linear Transport Theory.* New York: Addison-Wesley, 1967.

[102] A. Ishimaru, *Wave Propagation in Random Scattering Media.* New York: Academic Press, 1978.

[103] J.-M. Kaltenbach and M. Kaschke, "Frequency and time-domain modelling of light transport in random media," in *Medical Optical Tomography: Functional Imaging and Monitoring*, vol. IS 11, pp. 65–86, SPIE Press, 1993.

[104] W. Briggs, *A Multigrid Tutorial.* Philadelphia: Society for Industrial and Applied Mathematics, 1987.

[105] K. Lange, "Convergence of EM image reconstruction algorithms with Gibbs smoothing," *IEEE Trans. on Medical Imaging*, vol. 9, no. 4, pp. 439–446, December 1990.

[106] M. Abramowitz and I. Stegun, eds., *Handbook of Mathematical Functions.* New York: Dover, 1965.

APPENDICES

APPENDIX A: ML ESTIMATE OF DOSAGE

In this appendix, we derive the ML estimate of the dosage parameter y_T. For the following development, we assume our observations are the random integral projection measurements $\{Z_i = \ln(\frac{y_T}{Y_i})\}$. Consider the log-likelihood function of z in terms of the unknown dosage parameter y_T. Let the i-th actual discretized projection measurement across X be $\tilde{z}_i = A_{i*}x$. Note that Y_i are Poisson distributed with mean and variance $y_T e^{-\tilde{z}_i}$. Then from transformation of the Poisson distribution in (2.24), we have

$$\mathcal{P}(z|x) = \prod_{i=1}^{M} \frac{\exp\left\{-y_T e^{-\tilde{z}_i}\right\} \left(y_T e^{-\tilde{z}_i}\right)^{\left(y_T e^{-z_i}\right)}}{(y_T e^{-z_i})!} \tag{A.1}$$

for values of z_i corresponding to positive integer values of y_i and M denotes the number of projections. Stirling's formula provides a simplifying approximation for the factorial, which is relatively accurate for numbers in the typical range of transmission photon counts [106]:

$$\left(y_T e^{-z_i}\right)! \approx \left(2\pi y_T e^{-z_i}\right)^{\frac{1}{2}} \left(y_T e^{-z_i}\right)^{y_T e^{-z_i}} \exp\left\{-y_T e^{-z_i}\right\} \quad . \tag{A.2}$$

Substituting (A.2) into (A.1) and taking the logarithm we obtain

$$\log \mathcal{P}(z|x) = \sum_{i=1}^{M} \left\{ y_T \left(e^{-z_i} - e^{-A_{i*}x}\right) + y_T e^{-z_i} \left(z_i - A_{i*}x\right) - \frac{1}{2} \log\left(2\pi y_T e^{-z_i}\right) \right\} . \tag{A.3}$$

When x is known, the ML estimate of y_T can be obtained by differentiating (A.3) with respect to y_T and setting the result to zero.

$$\hat{y}_T = \frac{M}{2\sum_{i=1}^{M} \left[e^{-z_i} - e^{-A_{i*}x} + e^{-z_i}(z_i - A_{i*}x)\right]} \tag{A.4}$$

When x is unknown, the EM algorithm is used to compute the estimate. The update equation is obtained from (2.17) as

$$y_{T\,k+1} = \arg\max_{y_T} E[\log \mathcal{P}_{y_T}(z|X)|Z = z, y_{T\,k}, \sigma_k, p_k] \quad . \tag{A.5}$$

By substituting (A.3) into (A.5) and then maximizing with respect to y_T, we can obtain the update equation for y_T as

$$y_{T\,k+1} = \frac{M}{2\sum_{i=1}^{M}\left[e^{-z_i} - E\left[e^{-A_{i*}X} \mid Z = z,\ \sigma_k,\ p_k,\ y_{T\,k}\right] + e^{-z_i}(z_i - A_{i*}\bar{X})\right]} \quad , \tag{A.6}$$

where

$$\bar{X} = E\left[X \mid Z = z,\ \sigma_k,\ p_k,\ y_{T\,k}\right]$$

is the expected value of X given the projections and the model parameters at iteration k. The update equation (A.6) is exactly the same as (A.4) with the functions of unknown x replaced by their posterior expectations.

APPENDIX B: ML ESTIMATE OF THE SCALING PARAMETER λ_K

In this appendix, we outline the derivation of the update equations to compute the conditional ML estimate of λ_k defined in (4.27) and (4.28) using the EM algorithm. The EM algorithm [14, 15] hypothesizes the existence of a set of complete data from which ML estimation would be feasible, then attempts to iteratively maximize over the expectation of the complete data. In this case, the complete data at any scale k consists of the observations $y^{(l)}$, the SMAP estimate $\hat{x}^{(k+1)}$, and the unknown image $x^{(k)}$. The iterative set of update equations that results from the EM algorithm is given as

$$\hat{\lambda}_L(t+1) = \arg\max_{\lambda_L} E\left[\log P(X^{(L)}|\lambda_L)|y^{(l)}, \hat{\lambda}_L(t)\right] \tag{B.1}$$

$$\hat{\lambda}_k(t+1) = \arg\max_{\lambda_k} E\left[\log P(X^{(k)}|\hat{x}^{(k+1)}, \lambda_k)|y^{(l)}, \hat{x}^{(k+1)}, \hat{\lambda}_k(t)\right] \quad k \neq L, \tag{B.2}$$

where t denotes the discrete iteration index.

The log-likelihood of the image model at scale $k = L$, with explicit dependence of the partition function on the scale parameter (refer to Section 2.2.1), is given as

$$\log P(x^{(L)}|\lambda_L) = \frac{-1}{p\lambda_L^p} \sum_{\{i,j\}\in\mathcal{N}} b_{i-j} \left|x_i^{(L)} - x_j^{(L)}\right|^p - N_L \log \lambda_L + C \ , \tag{B.3}$$

where N_L is the number of pixels at scale $k = L$ and C is used to denote a constant independent of $x^{(L)}$ and λ_L. For scales $k \neq L$, the log-likelihood of the image model is given as

$$\log P(x^{(k)}|x^{(k+1)}, \lambda_k) = \frac{-1}{p\lambda_k^p} \sum_{\{i,j\}\in\mathcal{N}} \frac{b_{i-j}}{\hat{\sigma}_{ij}^p(x^{(k+1)})} \left|x_i^{(k)} - x_j^{(k)}\right|^p - N_k \log \lambda_k + C \ , \tag{B.4}$$

where N_k is the number of pixels at scale k and C is used to denote a constant independent of $x^{(k+1)}$ and λ_k. Substituting (B.3) and (B.4) into (B.1) and (B.2), we obtain for $k = L$

$$\hat{\lambda}_L(t+1) = \arg\max_{\lambda_L} \left\{ \frac{-1}{p\lambda_L^p} \sum_{\{i,j\}\in\mathcal{N}} b_{i-j}\, E\left[\left| X_i^{(L)} - X_j^{(L)} \right|^p |y^{(l)}, \hat{\lambda}_L(t) \right] - N_L \log \lambda_L \right\},$$

and for $k \neq L$

$$\hat{\lambda}_k(t+1) =$$
$$\arg\max_{\lambda_k} \left\{ \frac{-1}{p\lambda_k^p} \sum_{\{i,j\}\in\mathcal{N}} \frac{b_{i-j}}{\hat{\sigma}_{ij}^p(\hat{x}^{(k+1)})}\, E\left[\left| X_i^{(k)} - X_j^{(k)} \right|^p |y^{(l)}, \hat{x}^{(k+1)}, \hat{\lambda}_k(t) \right] - N_k \log \lambda_k \right\}.$$

Maximizing the above set of equations with respect to λ_k yields (4.29) and (4.30).