Εξομάλυνση με χωρικά μεταβαλλόμενες μικτές κατανομές στην τομογραφική ανακατασκευή εικόνων

Η ΜΕΤΑΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ ΕΞΕΙΔΙΚΕΥΣΗΣ

υποβάλλεται στην

ορισθείσα από την Γενική Συνέλευση Ειδικής Σύνθεσης

του Τμήματος Μηχανικών Η/Υ και Πληροφορικής Εξεταστική Επιτροπή

από την

Κατερίνα Παπαδημητρίου

ως μέρος των Υποχρεώσεων για τη λήψη του

ΜΕΤΑΠΤΥΧΙΑΚΟΥ ΔΙΠΛΩΜΑΤΟΣ ΣΤΗΝ ΠΛΗΡΟΦΟΡΙΚΗ ΜΕ ΕΞΕΙΔΙΚΕΥΣΗ ΣΤΙΣ ΤΕΧΝΟΛΟΓΙΕΣ - ΕΦΑΡΜΟΓΕΣ

Ιούνιος 2015

Acknowledgements

At this point I would like to express my gratitude to those whose contribution was instrumental in the completion of this work. I am truly grateful to my supervisor Prof. Christophoros Nikou for his trust, significant guidelines and continuous support during this period of time.

Finally, I would like to extend my appreciation to my family and my friends who are always by my side and support me.

Table of Contents

1	Introduction				
	1.1	Tomog	graphy	1	
	1.2	Types	of Tomography	2	
		1.2.1	Transmission Computed Tomography	2	
		1.2.2	Emission Computed Tomography	4	
	1.3	Struct	ure of the thesis	6	
2	Basi	c Princ	iples of Image Reconstruction	7	
	2.1	Introd	uction	7	
	2.2	The R	adon Transform	9	
	2.3	The Fe	ourier-Slice Theorem	11	
	2.4	Recons	struction by Filtered Backprojection	12	
3	Itera	ative Im	age Reconstruction	14	
	3.1	Introd	uction	14	
	3.2	Genera	al Concept of Iterative Methods	16	
	3.3	Statist	ical Model of Event Counts	18	
		3.3.1	The Poisson model	18	
		3.3.2	The Gaussian model	19	
	3.4	Image	reconstruction criteria	19	
		3.4.1	Maximum-Likelihood criterion	20	
		3.4.2	Maximum a Posteriori criterion	20	
	3.5	Iterati	ve Reconstruction Algorithms	22	
		3.5.1	The Maximum-Likelihood Expectation-Maximization Algorithm	24	
		3.5.2	Maximum A Posteriori Reconstruction Algorithms	27	
		3.5.3	Gibbs prior	29	
		3.5.4	Total Variation prior	30	
		3.5.5	Preconditioned Conjugate Gradient	31	
4	Tomographic image reconstruction with spatially varying mixture models				
	4.1	Introd	uction	33	
	4.2	Spatia	lly Varying Gaussian Mixture Models	35	
		4.2.1	Continuous, Gamma Distributed Line Process Model $\ .$	37	

		4.2.2	Binary, Bernoulli Distributed Line Process Model	40
	4.3	Spatia	lly Varying Mixture Models with Gamma Mixture Prior	41
		4.3.1	Continuous, Gamma Distributed Line Process Model	42
		4.3.2	Binary, Bernoulli Distributed Line Process Model	43
	4.4	Exper	imental results	43
5	Con	clusions	and future work	60

List of Figures

1.1	Compton Scattering.	3
1.2	(a) First Generation CT. (b) Second Generation CT	4
1.3	(a) Third Generation CT. (b) Fourth Generation CT	4
2.1	(a) Two objects, an input parallel beam and a detector. (b) The result of back-	
2.2	projecting the 1-D absorption profile. \dots (b) The intensity of the two individual	8
	back-projections. \ldots	8
2.3	Reconstruction using an increasing number of backprojections. The original ob-	
	ject is shown at the bottom.	9
2.4	Geometry of the line integrals associated with the Radon transform	10
2.5	The normal parameters that stipulate the position of the line	10
2.6	(a) The Shepp-Logan phantom and (b) its sinogram	11
2.7	The Fourier slice theorem.	12
2.8	Shepp-Logan phantom reconstructed by (a) non filtered backprojection and (b)	
	filtered backprojection using a Hamming window.	13
3.1	A general model of tomographic projection in which the measurements are given	
	by weighted integrals of the emitting object distribution. $[12]$	16
3.2	A discrete model of the projection process	16
3.3	Illustration of a single element of the system model \mathbf{H}	18
3.4	Comparison of objective functions in a simple 1D example [12]	22
3.5	Flowchart of a generic iterative reconstruction algorithm [12]	24
3.6	The maximum-likelihood expectation-maximization algorithm in the form of the	
	general iterative model [12]. \ldots \ldots \ldots \ldots \ldots \ldots	27
4.1	First-order neighborhood cliques in the contextual mixing proportions mesh. (a)	
	Each MRF site is associated with a probability scalar value π_j^n and is dependent	
	on 4 neighbors. (b) The set of horizontal neighbors, $\gamma_1(n)$, is highlighted. (c)	
	The set of vertical neighbors, $\gamma_2(n)$, is highlighted [32]	36
4.2	The Student's $t\text{-}distribution$ for various degrees of freedom	38

4.3	Graphical model for the edge preserving model. Superscripts and subscripts	
	$n, k \in [1, N]$ denote pixel index, subscript $j \in [1, K]$ denotes segment index,	
	$d \in [1,D]$ describes the neighborhood direction type. Γ equals the maximum	
	number of possible neighbors. The Figure is reproduced by [27]	39
4.4	Graphical model for the binary line process edge preserving model. Superscripts	
	$n, k \in [1, N]$ denotes pixel index, subscript $j \in [1, J]$ denotes kernel (segment)	
	index, $d \in [1, D]$ describes the neighborhood direction type and $l \in [1, \Gamma]$ denotes	
	neighbor index. The Figure is reproduced from [27]	41
4.5	Graphical representation of the proposed model	42
4.6	Graphical model for the binary line process edge preserving model	43
4.7	(a) Shepp - Logan phantom. (b) Elliptical phantom	44
4.8	Comparative statistics for various performance indices for the Shepp-Logan	
	phantom for 75 photons per pixel. (a) ISNR (mean values of the 40 experi-	
	ments), (b) structural similarity (mean value), (c) bias, (d) variance	46
4.9	Comparative statistics for various performance indices for the Shepp-Logan	
	phantom for 55 photons per pixel. (a) ISNR (mean values of the 40 experi-	
	ments), (b) structural similarity (mean value), (c) bias, (d) variance. \ldots \ldots	47
4.10	Comparative statistics for various performance indices for the Shepp-Logan	
	phantom for 35 photons per pixel. (a) ISNR (mean values of the 40 experi-	
	ments), (b) structural similarity (mean value), (c) bias, (d) variance. \ldots .	48
4.11	Comparative statistics for various performance indices for the Shepp-Logan	
	phantom for 15 photons per pixel. (a) ISNR (mean values of the 40 experi-	
	ments), (b) structural similarity (mean value), (c) bias, (d) variance. \ldots \ldots	49
4.12	eq:comparative statistics for various performance indices for the Elliptical phantom	
	for 80 photons per pixel. (a) ISNR (mean values of the 40 experiments), (b)	
	structural similarity (mean value), (c) bias, (d) variance	51
4.13	Comparative statistics for various performance indices for the Elliptical phantom	
	for 56 photons per pixel. (a) ISNR (mean values of the 40 experiments), (b)	
	structural similarity (mean value), (c) bias, (d) variance	52
4.14	Comparative statistics for various performance indices for the Elliptical phantom	
	for 36 photons per pixel. (a) ISNR (mean values of the 40 experiments), (b)	
	structural similarity (mean value), (c) bias, (d) variance	53
4.15	Comparative statistics for various performance indices for the Elliptical phantom	
	for 24 photons per pixel. (a) ISNR (mean values of the 40 experiments), (b)	
	structural similarity (mean value), (c) bias, (d) variance	54
4.16	The estimated images for the Shepp-Logan phantom with 75 photons per pixel.	55
4.17	The estimated images for the Shepp-Logan phantom with 55 photons per pixel.	55
4.18	The estimated images for the Shepp-Logan phantom with 35 photons per pixel.	55
4.19	The estimated images for the Shepp-Logan phantom with 15 photons per pixel.	56
4.20	The estimated images for the Elliptical phantom with 80 photons per pixel	56
4.21	The estimated images for the Elliptical phantom with 56 photons per pixel	56

4.22	The estimated images for the Elliptical phantom with 36 photons per pixel	57
4.23	The estimated images for the Elliptical phantom with 24 photons per pixel	57
4.24	Comparison of horizontal profiles between the original Shepp-Logan phantom	
	and the reconstructed images provided by the proposed GAMMA-CLP (PCG)	
	and the Gibbs prior for 75 counts per pixel. \ldots \ldots \ldots \ldots \ldots \ldots	58
4.25	Comparison of horizontal profiles between the original Elliptical phantom and	
	the reconstructed images provided by the proposed GAMMA-CLP (PCG) and	
	the Gibbs prior for 75 counts per pixel. \ldots \ldots \ldots \ldots \ldots \ldots \ldots	58
4.26	Execution times for the compared algorithms (60 iterations, $\epsilon = 10^{-3}$)	59

Abstract

Papadimitriou Katerina.

MSc, Computer Science Department, University of Ioannina, Greece. June, 2015. Tomographic image reconstruction with spatially varying mixture models.

Thesis Supervisor: Christophoros Nikou.

Image reconstruction is a mathematical process through which an accurate image of an object is reconstructed from multiple projections. Tomographic reconstruction strategies have gained tremendous attention in the last decades, due to their crucial role in non-invasive visualization of the interior of objects such as the human body. Applications of these methods are, among others, radiology, geophysics and material science.

Although, several approaches have been proposed for solving the tomographic reconstruction problem, such as analytical reconstruction techniques, the main research tendencies are centered on iterative image reconstruction. This approximation inevitably requires repeated projection and back projection procedures, signifying that the estimated image is progressively refined in a repetitive calculation.

The basic problem of tomographic reconstruction is the estimation of the attenuation coefficient, which leads to noisy data. This assumption is based on the knowledge that, because of its blurring effect, the system (projection) matrix suppresses image detail. Therefore any such detail present in the reconstructed image is more probably to have been caused from noise. A regular method for addressing the problem of noise propagation is the Bayesian maximum a posteriori (MAP) algorithm, which imposes a prior probability on the image to be reconstructed and usually aims to encourage the image to be smooth, so as to suppress the effect of noise.

The purpose of this study is the effective noise elimination and the preservation of region boundaries in the reconstructed image. To this end, we present four models which are based on maximum a posteriori estimation and use two different priors: a Gaussian mixture prior and a Gamma mixture prior. Simultaneously, in order to account for the modeling of edges between image segments, appropriate MRF smoothness priors, which are based on Student's **t**-distribution and Bernoulli distributions formalized as a line process, on the contextual mixing proportions are employed, which model the existence or not of boundaries. The overall algorithm consists of two alternating steps. At first, the mixture model parameters are automatically estimated from the image and then the

reconstructed image is estimated by optimizing the MAP criterion using the one-steplate-EM (OSL-EM) or the preconditioned conjugate gradient (PCG) algorithms.

Περίληψη

Παπαδημητρίου Κατερίνα. MSc, Τμήμα Πληροφορικής, Πανεπιστήμιο Ιωαννίνων Ιούνιος, 2015. Εξομάλυνση με χωρικά μεταβαλλόμενες μικτές κατανομές στην τομογραφική ανακατασκευή εικόνων Επιβλέπων καθηγητής: Χριστόφορος Νίκου.

Ένας θεμελιώδης τομέας στην επεξεργασία εικόνας σχετίζεται με το πρόβλημα της ανακατασκευής εικόνας από προβολές. Ο όρος ανακατασκευή εικόνας εκφράζει τη μαθηματική διαδικασία, μέσω της οποίας προσπαθούμε να ανακατασκευάσουμε μια καθαρή εικόνα ενός αντικειμένου από πολλαπλές προβολές, δηλαδή τις τιμές της έντασης της εξασθενημένης ακτινοβολίας που εξέρχεται από το αντικείμενο. Οι μέθοδοι ανακατασκευής εικόνας έχουν γίνει αρκετά δημοφιλείς κατά τις τελευταίες δεκαετίες λόγω του κρίσιμου ρόλου τους στη μη επεμβατική απεικόνιση του εσωτερικού των αντικειμένων, όπως το ανθρώπινο σώμα. Εφαρμογές των μεθόδων αυτών είναι, μεταξύ άλλων, η ακτινολογία, η γεωφυσική και η επιστήμη των υλικών.

Αν και έχουν προταθεί διάφορες προσεγγίσεις για την επίλυση του προβλήματος της τομογραφικής ανακατασκευής, όπως είναι οι στρατηγικές αναλυτικής ανακατασκευής, οι κύριες ερευνητικές τάσεις επικεντρώνονται στις επαναληπτικές μεθόδους ανακατασκευής. Αυτή η προσέγγιση απαιτεί επαναλαμβανόμενες διαδικασίες προβολής και οπισθοπροβολής για την εκτίμηση της εικόνας.

Το βασικό πρόβλημα της τομογραφικής ανακατασκευής είναι ο υπολογισμός του συντελεστή εξασθένησης που οδηγεί σε θορυβώδη δεδομένα. Μια διαδεδομένη επαναληπτική μέθοδος για την αντιμετώπιση του προβλήματος της διάδοσης του θορύβου είναι η μεγιστοποίηση της εκ των υστέρων πιθανοφάνειας (MAP). Αύτη η τεχνική στοχεύει στην εξομάλυνση της εικόνας καταστέλλοντας την επίδραση του θορύβου.

Ο σχοπός της παρούσας μελέτης είναι η αποτελεσματική εξάλειψη του θορύβου και η διατήρηση των αχμών στην ανακατασκευασμένη εικόνα. Για το σχοπό αυτό, προτείνουμε τέσσερα νέα μοντέλα τα οποία βασίζονται στη μέγιστη εκ των υστέρων εκτίμηση (MAP) και χρησιμοποιούν δύο διαφορετικές εκ των προτέρων πιθανότητες: μια Γκαουσιανή μικτή κατανομή και μια Γάμμα μικτή κατανομή. Παράλληλα, προκειμένου να μοντελοποιηθούν οι ακμές μεταξύ των περιοχών της εικόνας, κατάλληλες εκ των προτέρων πιθανότητες βασισμένες σε τυχαία πεδία Markov (MRF), μια Student's **t** και μια Bernoulli κατανομή, εφαρμόζονται στα βάρη των συνιστωσών της μικτής κατανομής, δηλαδή στα διανύσματα πιθανότητας. Η κύρια συνεισφορά αυτής της εργασίας είναι η αποτελεσματικότητα των προτεινόμενων εκ των προτέρων πιθανοτήτων, οι οποίες παρέχουν μια ομαλά ανακατασκευασμένη εικόνα διατηρώντας τη δομή των ακμών. Επίσης, όλες οι παράμετροι των μοντέλων υπολογίζονται αυτόματα από τα δεδομένα.

Chapter 1

Introduction

1.1 Tomography

1.2 Types of Tomography

1.3 Structure of the thesis

1.1 Tomography

Medical imaging comprises different imaging modalities and processes to view human body for diagnostic, monitoring and treatment purposes, and therefore has a crucial role in the amelioration of healthcare. The progress in computer technology leads to the evolution of improved medical imaging equipment, so that more accurate information can be captured with non invasive methods.

A prominent technique for human internal structure imaging is tomography. Tomography (Greek tomos= section, + graphein= record) ascribes the cross-sectional imaging of a 3-D region of human body from either transmission or emission data, obtained by impinging penetrating waves on the object from many different directions. The influence of this method in diagnostic medicine has been unprecedented, since it has enabled doctors to view internal organs with exceptional precision and safety to the patient [1]. Therefore, except from medicine, tomography has found extensive application in many scientific fields, including radiology, archaeology, biology, geophysics, astrophysics and others.

Tomographic imaging variations provide different kind of information depending on the different types of data acquisition. Computed Tomography (CT) uses special X-ray equipment to generate imaging of a cross section of the human body, while magnetic resonance imaging (MRI) attains imaging of the internal body through magnetic fields and radio waves. On the other hand, Single Photon Emission Computed Tomography (SPECT) is a nuclear medicine tomographic imaging technique utilizing Gamma rays and Positron Emission Tomography (PET) is based on the detection of positrons, emitted by a radionuclide in the region being examined. Optical tomography produces images from light, which is transmitted and scattered through the body. Images of the conductivity or permittivity of part of the body are gained through Electrical Impedance tomography and are inferred from surface electrode measurements.

In general, tomography is based on the mathematical formula that was first expressed by Radon in 1917 [3]. Computed tomography can be separated into two main categories: a) the emission computed tomography and b) the transmission computed tomography. The main difference between these two methods is the source of the photons. In emission CT, the photon source is the object of interest, while in transmission CT the photon source is outside the object. Tomographic imaging involves image reconstruction from its projections, namely sinogram, computationally. Fundamentally, when radiation is transmitted on the human body, a part of the energy penetrates the body. Then, using a strip of absorption detectors on the other side, the amount of radiation is measured and transferred to a computer system. A sophisticated computer system, in turn, calculates and analyses data from each detector, and finally estimates and reconstructs multiple or two-dimensional cross-sectional images.

1.2 Types of Tomography

Although, the mathematical solution to the reconstruction problem was initially expressed by Radon in 1917, the revolution in tomographic imaging originated in 1972 with Hounsfield's invention of the first Computed Tomography (CT) scanner, using a computer to produce cross-sectional images (slices) of the human brain through projection data. SPECT and PET methods are the most popular imaging technologies of this category. In transmission CT the goal is to measure the linear attenuation coefficients of the internal organs using X-rays.

1.2.1 Transmission Computed Tomography

Computed tomography uses X-ray transmission measurements to generate a slice of the human body. This is the most widely used type of CT and is currently one of the primary applications in digital image processing in medicine. The X-rays are able to pierce the body of the patient. However, the entire amount of radiation does not pass through the body because a proportion of energy is dispersed within the body losing part of it. During this scattering, a photon interacts with an electron into the body, transfer part of the energy in it, and displaces the electron. X-ray then bounces in a new direction with diminished energy. This is called Compton scattering and shown in Figure 1.1.

Another part of the energy vanishes in the body, altering the energy of the tissues of the body. Radiation release into the human body may cause damage if the radiation quantity is high. Assume that the intensity of the radiation before entering the object is



Figure 1.1: Compton Scattering.

 I_0 , and the intensity after leaving the object is I_d . Then, the two intensities are related by:

$$\frac{I_d}{I_0} = exp(-p) \tag{1.1}$$

where p is the contour integral of linear attenuation coefficient along X-rays and is given by:

$$p = \ln \frac{I_0}{I_d} \tag{1.2}$$

The aim of tomography is to create an image with different linear attenuation coefficients within the body. The attenuation coefficient, which is denoted by μ , is the absorption's proportion of each material. Bones have higher values of μ , while soft tissues have lower and air regions have even lower value. That's the reason why the tomographic imaging is introduced in gray scale, which corresponds to escalations of attenuation coefficients. The highest value corresponds to white, like bones and lowest in black.

First generation CT employs a pencil X-ray beam and a single detector as shown in Figure 1.2 (a). This source-detector pair counts parallel projections, one sample at a time, by stepping linearly across the object. After each projection, the source-detector assembly rotates to a new position for the next projection [4].

The system was slow, however, with typical acquisition times of 4 min per section, even for relatively low-resolution images [5].

Second Generation CT scanners act on the same principle as G1 scanners, however, the beam used is in the form of a fan (Figure 1.2 (b)). This lets the use of multiple detectors, thus demands fewer transportations of the source-detector pair.

Third Generation scanners are a significant enhancement over the earlier two generations of CT geometries. As Figure 1.3 (a) shows, G3 scanners use a bank of detectors in order to cover the entire field of view of a beam. Consequently, each increase of angle



Figure 1.2: (a) First Generation CT. (b) Second Generation CT.

generates a whole projection, erasing the requirement of translating the source-detector combination, as in geometry of G1 and G2 scanners.



Figure 1.3: (a) Third Generation CT. (b) Fourth Generation CT.

Fourth Generation scanners (Figure 1.3 (b)) move a step beyond. By utilizing a circular ring of detectors only the source has to spin. The advantage of G4 is the high speed but the need of greater X-ray scatters, which lead to more cost, is the main drawback of these scanners.

1.2.2 Emission Computed Tomography

Emission CT utilizes decay of radioactive isotopes to illustrate the distribution of the isotope as a function of time. These isotopes may be medicated to the patient in the form of radiopharmaceuticals either by injection or by aspiration. Thus, by administering a

radioactive isotope, the path of the isotope can be traced through the entire of the body.

Radioactive isotopes are characterized by the emission of Gamma-ray photons or positrons, both products of nuclear decay. The concentration of such an isotope, in any cross section, changed over time due to radioactive decay, flow, and biochemical kinetics within the body. This entails that all the data for one cross-section must be gathered in a time interval that is short, compared to the time constant associated with the changing concentration. This aspect provides emission CT with great capabilities and utility in diagnostic medicine, because now by analysing the images taken at different times for the same cross section the functional state of various organs in a patient's body can be determined [1]. Emission CT consists of two kinds: single photon emission CT and positron emission CT.

Single photon emission Computed Tomography (SPECT)

Single photon emission CT employs Gamma camera, which obtains multiple 2D images from various angles. Subsequently, it is pursued by the tomographic reconstruction procedure, accomplishing a 3D representation. These 3D data can be managed accordingly in order to acquire thin sections along any desirable axis.

The principal difference between SPECT and PET scan is that in the case of SPECT Gamma radiation is emitted and estimated straightforward, while in PET there are positron emission decays into two photons.

More specifically, the Gamma camera rotates around the patient, and simultaneously, gets views at intervals of 3 to 6 degrees. The time required to achieve each projection is 15 to 20 seconds so the total time amounts to 15 to 20 minutes. Nonetheless, multiple Gamma cameras can decrease the execution time [1].

Positron emission Computed Tomography (PET)

To conduct the scan, an isotope enters the body of the patient. There is a waiting interval until the gist to be expanded across the tissues of body, and afterwards, the patient is placed in the imaging scanner. The scanning process takes at most 1 hour and during this period the concentrations of the substance in different tissues are recorded.

The isotopes that are used in PET are unlike these used in other positron emission modalities such as SPECT. Most of them are short-lived and their selection is based on the organ under investigation. The radiopharmaceuticals have radionuclides, which are positron transmitters. Positron is an elementary particle, which has the same mass as an electron and has a positive electrical charge. When the positron is emitted, it interacts with surrounding tissue and finally is devastated with an electron, resulting in energy loss. The result of annihilation is the production of a pair of photons of Gamma radiation, which run at about opposite directions. The photons are detected when they reach a scintillator the scanning device, causing a burst of light detected by photomultiplier tubes or silicon avalanche photodiodes [1].

1.3 Structure of the thesis

Filtered back-projection is a direct algorithm that is mainstream in the reconstruction problem, both because of its speed and its accuracy. The projection data collected are associated with the Fourier transform of the object, which is illustrated through the Fourier slice theorem. Moreover, based on this theorem, the 2D inverse Fourier transform can be forged to generate the filtered back-projection (FBP) algorithm [1]. Filtered Back Projection is presented in Chapter 2, including the Radon transform which constitutes the foundation of FBP technique.

In the early 1980's, new methods of image reconstruction in emission tomography were proposed, which could overcome some drawbacks of the aforementioned conventional methods. These methods could take into account physical traits of emission tomography, such as the Poisson nature of photons, and many other factors leading to better outcomes, by combining these with multiple iterations of reconstruction. For that reason they are termed as statistical iterative reconstruction algorithms. Such iterative reconstruction techniques may result in more execution time but, also, in substantially less noisy images through complex modeling of detector response and of the statistical behavior of measurements. Therefore, the most dominant iterative algorithms which are based on statistical estimation criteria are presented in Chapter 3.

In Chapter 4 we focus on one of these iterative reconstruction techniques called Maximum A Posteriori Expectation Maximum (MAP-EM) reconstruction. MAP-EM with a Gibbs prior has proven to produce more precise images than filtered back projection algorithms. We propose four different models, which are called spatially varying mixture models, and are carried out through a Gaussian mixture and a Gamma mixture prior. Concurrently, in order to model the existence of boundaries, appropriate MRF smoothness priors, which are based on Student's-t and Bernoulli distributions formalized as a *line process*, on the contextual mixing proportions are employed. These formulas may not only potentially reduce the noise in the system, due to photon scattering, nevertheless they preserve region boundaries. Finally, a discussion of these new models and the experimental results are detailed.

Conclusions and an outline of possible future work is given in Chapter 5.

Chapter 2

Basic Principles of Image Reconstruction

2.1 Introduction

2.2 The Radon Transform

- 2.3 The Fourier-Slice Theorem
- 2.4 Reconstruction by Filtered Backprojection

2.1 Introduction

Image reconstruction is a mathematical procedure that generates images from ray projection data, acquired at many different angles around the region of interest. This process has a decisive influence on image quality and, therefore, it is significant to reconstruct images with the lowest possible noise without sacrificing image accuracy and spatial resolution.

The problem of reconstructing is important and can be explained in a straightforward natural manner. To begin, consider Fig. 2.1(a) which is composed of two objects on a uniform background. Consider it as a cross section of a region of the human body. Suppose that a thin, flat beam of rays traverses from left to right and that the energy of the beam is absorbed more by the objects than by the background. Then, the detector on the other side provides the signal shown, which is proportional to absorption. Any point in the signal is the sum of the absorption values across the single ray in the beam corresponding spatially to that point. The reconstruction begins by creating an image based on just this information, thus, the 1-D signal is projected back across the direction from which the beam came, as Fig. 2.1(b) shows. The process of back-projecting a 1-D signal across a 2-D area means duplicating the same 1-D signal across the image to the direction of the beam.

Next, suppose that the position of source-detector pair is rotated by 90° , as in Fig. 2.2(a). Repeating the previous operation, a back-projected image in the vertical direction



Figure 2.1: (a) Two objects, an input parallel beam and a detector. (b) The result of back-projecting the 1-D absorption profile.

is accomplished. Fig. 2.2(b) arises from adding the two back-projections. It is unambiguous that the objects of interest are included in the squares shown. However, the shape of



Figure 2.2: (a) The beam and detector rotated by 90° . (b) The intensity of the two individual back-projections.

the objects is not obvious, so, by taking more views in the way just presented the shape is revealed. As the number of projections increases, brighter regions dominate the result and back-projections with few or no intersections fade into the background [4]. Figure 2.3 illustrates this concept. The approach just described is called *backprojection*. Although the reconstructed image is a reasonably good approximation to the shape of the original object, the resulting image is blurred. Blurring issue in image reconstruction is crucial and its solution is addressed in Section 2.4.

The goal of image reconstruction is to use the data gathered by the detectors in order to frame the image of the object. There are two major categories of reconstruction methods. The first is the analytic and is based on the Fourier transform of the object and its projections. Analytical reconstruction is currently extensively used on clinical CT scanners because of their computational effectiveness and numerical constancy. Many methods based on this approach have been elaborated for different generations of CT data procuring geometries, from axial parallel-beam and fan-beam CT to current multi-slice helical and cone-beam CT.

This Chapter will briefly deliberate how this reconstruction is conducted. Firstly, the Radon transform is introduced which delimits the relationship between the object and



Figure 2.3: Reconstruction using an increasing number of backprojections. The original object is shown at the bottom.

its projections. Then, the Fourier slice theorem will be presented. It is shown that this theorem is essential to the Filtered Back Projection (FBP) method, which reconstructs the object from its projections. The second approach is the iterative methods that model the data collection process and attempt to find the image that is most compatible with the measured data. The iterative algorithms will be analysed in Chapter 3.

2.2 The Radon Transform

The foundation of analytical reconstruction methods is the Radon transform [2], which relates a 2D function f(x, y) to the set of line integrals of that function. For now on, only projections of two-dimensional objects will be considered. Assume that parallel beams are used and that a ray penetrates the object, which is represented by a function f(x, y), like Figure 2.4 illustrates. The parameter R is the shortest distance from the origin of the coordinate system to the ray, and θ is an angle corresponding to the angular orientation of the ray. The significance of the normal parameters that stipulate the position of the ray are shown in Fig. 2.5.

The ray can be parametrized as $x \cos \theta + y \sin \theta = R$ and, in general, the projection $g_{\theta}(R)$ of a function f(x, y) is the line integral of the values of f(x, y) along the line inclined at an angle θ from the x-axis at a distance R from the origin:

$$g_{\theta}(R) = \int_{R \in line} f(x, y) dR \tag{2.1}$$

More precisely, an arbitrary point in the projection signal is given by the ray-sum along the line $x \cos \theta + y \sin \theta = R$. Because an integral is basically a sum of values, the value $g_{\theta}(R)$ is the sum of the values f(x, y) along the line in the *xy*-plane. For this reason, $g_{\theta}(R)$ is called a ray-sum and is given by:

$$g_{\theta}(R) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \delta(x \cos \theta + y \sin \theta - R) dx dy.$$
(2.2)



Figure 2.4: Geometry of the line integrals associated with the Radon transform.



Figure 2.5: The normal parameters that stipulate the position of the line.

where $\delta(\cdot)$ is the Dirac delta function.

This equation, which gives the projection (line integral) of f(x, y) along an impulsive line in the xy-plane is called as the Radon transform of the function f(x, y), on account of Johann Radon who proposed the formula in 1917 [3].

Consequently, by applying the Radon transform on an image f(x, y) for a given set of angles, the projection of the image along the given angles could be computed. The resulting projection is the sum of the intensities of the pixels in each direction and is a new image $g_{\theta}(R)$. On account of this, in the discrete case (2.2) becomes:

$$g_{\theta}(R) = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x,y) \delta(x\cos\theta + y\sin\theta - R) dxdy$$
(2.3)

with x, y, R and θ now being discrete variables. If θ is fixed and R varies, then (2.3) aggregates the pixels of f(x, y) along the line determined by the particular values of these two parameters. Reoccurring for all values of R engenders one projection. Changing θ and repeating the foregoing transaction accrues another projection, and so forth. This is precisely how the projections in Section 2.1 were generated. [4]

When the Radon transform, $g_{\theta}(R)$, is displayed as an image with R and θ being rectilinear coordinates, the result is called *sinogram*. A sinogram contains the data necessary to reconstruct f(x,y). To get a feeling of what the Radon transform is, the transform is applied to an example image. Figure 2.6(b) shows the Radon transform of the *Shepp – Logan head phantom*, this phantom is used throughout all experiments.



Figure 2.6: (a) The Shepp-Logan phantom and (b) its sinogram.

2.3 The Fourier-Slice Theorem

The aim of image reconstruction is to find the 2-D object f(x, y) given the set of 1-D projection signals, $g_{\theta}(R)$, obtained at various projection angles. If the 1-D Fourier transform of a projection with respect to R is:

$$G_{\theta}(\omega) = \int_{-\infty}^{\infty} g_{\theta}(R) e^{-j2\pi\omega R} dR \qquad (2.4)$$

with ω being the frequency variable, then substituting (2.2), the expression as emerges is:

$$G_{\theta}(\omega) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) e^{-j2\pi\omega(x\cos\theta + y\sin\theta)} dx dy.$$
(2.5)

By letting $u = \omega \cos \theta$ and $v = \omega \sin \theta$, the equation 2.5 changes into:

$$G_{\theta}(\omega) = \left[\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) e^{-j2\pi(ux + vy)} dx dy \right]_{u = \omega \cos \theta; v = \omega \sin \theta}.$$
 (2.6)

This expression is recognized to be equal to $F(x \cos \theta; y \sin \theta)$ where F(u; v) is the 2-D Fourier transform of the original 2-D function f(x; y):

$$F(u,v) = \int_{-\infty}^{\infty} f(x,y)e^{-j2\pi(ux+vy)}dxdy.$$
(2.7)

Thence, (2.6) turns into:

$$G_{\theta}(\omega) = [F(u, v)]_{u=\omega\cos\theta; v=\omega\sin\theta}.$$
(2.8)



Figure 2.7: The Fourier slice theorem.

Equation (2.8) is the *Fourier slice theorem* in image reconstruction from projections which is the basis of image reconstruction. The Fourier slice theorem declares that the 1-D Fourier transform is equal to a slice of the 2-D Fourier transform of the image estimated along a radial profile at angle θ with respect to the x-axis. This is schematically shown at Figure 2.7.

2.4 Reconstruction by Filtered Backprojection

As mentioned in Section 2.1, acquiring back-projections directly, returns blurred images. This problem may be solved by filtering the projections before computing the backprojections. The filtered back-projection method (FBP) relies on the result of the Fourier slice theorem. In this manner, using this result one can construct the two-dimensional Fourier transform of the object with the parallel projection data. Implementing a twodimensional inverse Fourier transform would then yield the original object.

FBP method can arise by writing the previous equations in a different way. We can write the 2-D inverse Fourier transform of the image f(x, y) in polar coordinates as:

$$f(x,y) = \int_0^{2\pi} \int_{-\infty}^{\infty} F(\omega\cos\theta, \omega\sin\theta) e^{j2\pi\omega(x\cos\theta + y\sin\theta)} \omega d\omega d\theta.$$
(2.9)

Then, using the Fourier Slice Theorem and the property [1] $G(\omega, \theta + \pi) = G(-\omega, \theta)$, the 1-D Fourier transform of the projection at angle θ is transmuted into:

$$f(x,y) = \int_0^\pi \int_{-\infty}^\infty |\omega| G_\theta(\omega) e^{j2\pi\omega(x\cos\theta + y\sin\theta)} d\omega d\theta =$$
(2.10)

$$\int_{0}^{\pi} \left[\int_{-\infty}^{\infty} |\omega| \, G_{\theta}(\omega) e^{j2\pi\omega R} d\omega \right]_{R=x\cos\theta + y\sin\theta}$$
(2.11)

The second equation is owed to the fact that the term $x \cos \theta + y \sin \theta$ is the constant R.

Equation (2.11) represents the estimation of the object f(x, y) using the Fourier transform of the projection data. The inner expression of the equation denotes a one-

dimensional filter function, with $|\omega|$ being a *ramp* filter and, for that reason, the image f(x; y) can be found by first filtering the projections with a ramp filter, and then integrating these filtered projection values at the coordinate $(x\cos\theta + y\sin\theta)$ over all projection angles θ . [4]

Summarizing the implementation of filtered backprojection is:

- Get the 1D Fourier transform of each projection.
- Multiply it by the filter function $|\omega|$.
- Obtain the inverse 1-D Fourier transform.
- Integrate all inverse transforms.

The filtered back-projection method (FBP) is a deterministic method that is based on the Fourier slice theorem and, thereby, on the inverse Radon transform. In order to retrieve a reconstruction of the object, a transfiguration of the measurements in the frequency domain back to the spatial domain is required. The FBP method occurs the interpolation in the spatial domain and these transformations are then multiplied by a weighting function. This approach is widely used in clinics due to the computational simplicity and convergence speed. On the other hand, FBP supplies suboptimal image quality, since the noise in the data is ignored. A resolution to this issue may be a reduction at the cut-off frequency of the ramp filter $|\omega|$, which diminishes the amount of noise in the resulted images, though leads to loss of resolution. Figure 2.8 shows the benefits of FPB without and with a Hamming window.



Figure 2.8: Shepp-Logan phantom reconstructed by (a) non filtered backprojection and (b) filtered backprojection using a Hamming window.

Chapter 3

Iterative Image Reconstruction

3.1 Introduction

- 3.2 General Concept of Iterative Methods
- 3.3 Statistical Model of Event Counts
- 3.4 Image reconstruction criteria
- 3.5 Iterative Reconstruction Algorithms

3.1 Introduction

Despite the efficiency and elegance of analytical approaches, the precision of the resulting images is restricted by the approximations of the reconstruction formula. In contrast, iterative methods, which are elaborated in this chapter, can assimilate explicitly the mapping between the source and the sinograms. As mentioned, another deficiency of the analytic techniques is that they do not take into consideration the statistical variability, that pertains to photon inadequate detection. Even though, the resulting noise can be restrained by diversifying the cut-off frequency of the ramp filter, this type of filtering is less sufficient in effectuating an optimal trade-off, since the noise is signal reliant. On the other hand, momentous physical factors can be more accurately incorporated into iterative reconstruction, rendering lower image noise and higher spatial resolution [7].

The iterative reconstruction (IR) methods include statistical reconstruction (SIR) algorithms and algebraic reconstruction techniques (ART). ART reckon that the cross section involves an array of unknowns and a system of linear equations have to be solved iteratively, in terms of the measured projection data. Unfortunately, ART are less distinguished than the statistical ones in the tomographic field. In the late 1990's, SIR methods were initially presented commercially for reconstructing purposes of SPECT and PET, after many years of academic research, replacing FBP algorithm clinically for PET and SPECT. Complete SIR methods for X-ray CT was dispensable in about 2011, intending to limit patient X-ray dose. Latest clinical surveys on early versions of iterative reconstruction indicated a potential dose diminution of up to 65 % compared with FBP-based reconstruction algorithms.

Iterative reconstruction techniques adopt a generalized linear model, that enables the integration of essential corrections for image degrading effects, such as attenuation, scatter and depth-dependent resolution. Specifically, IR methods associate the unknown object with the ideal measurements in the absence of noise. This is obtained by incorporating probabilistic models of the noise and, in the case of Bayesian methods, of the image itself. The price of these enhancements is the high computational effort, since the analytic form of the solution is impractical to compute. Hence, most reconstruction algorithms that attempt to incorporate a precise imaging model are iterative, namely, the object is progressively estimated through a repetitive calculation, which leads to excessive processing time. In the beginning, this shortcoming prevented the transition of iterative techniques from the research lab to the clinic, however, now are widely accepted, owing to the advances in computer speed and the development of efficient modeling techniques and fast reconstruction algorithms.

Iterative algorithms require repeating executions of projection and back-projection processes. More precisely, projection data are evaluated based on an initial attenuation coefficients of all pixels. Then, the projections corresponding to the current estimate are compared with the measured projections, and the pixel attenuation values are revised until a desired level of error between the predicted and measured data is obtained.

Firstly, a classic image reconstruction mechanism performed in tomography, is the maximum likelihood expectation maximization (MLEM) algorithm. However, major drawbacks of MLEM algorithm such as the slow convergence and high computational cost enforced Hudson and Larkin to elaborate the ordered subsets expectation maximization (OSEM) algorithm [10]. The OSEM algorithm is an efficient variation of the MLEM which intended to diminish reconstruction time and cost and to facilitate clinical use [10]. Later, Browne and De Pierro, in order to enhance the convergence, developed the row action maximum likelihood algorithm (RAMLA) [11].

A regular method for manipulating the problem of slow convergence rate and noise propagation is the Bayesian maximum a posteriori (MAP) algorithm which imposes a priori information as a regularization constraint and amplifies the spatial smoothness of reconstructed image. A wide range of priors are reviewed and proposed in the literature. Particularly, a popular Bayesian prior is the Gibbs distribution, whose computational process relies on the local differences between neighboring pixels. Another prior is the Total Variation, which suppresses unstable oscillations while preserving object edges. This chapter presents the general principles of iterative image reconstruction and a classification of iterative methods into a few major types.

3.2 General Concept of Iterative Methods

A tomographic reconstruction problem can be framed as follows: Find the object distribution \mathbf{f} , given a set of projection measurements \mathbf{g} and the system matrix \mathbf{H} , which describes the relationship between object \mathbf{f} and the projection \mathbf{g} (Fig. 3.1).



Figure 3.1: A general model of tomographic projection in which the measurements are given by weighted integrals of the emitting object distribution. [12]

Without loss of generality, in this review the problem of two-dimensional image reconstruction is described. For computing purposes, the reconstructed image cannot be represented by a continuous domain function; alternatively, a sampled version of the image, in a discrete domain, is estimated (Fig.3.2).



Figure 3.2: A discrete model of the projection process.

Thus, by using vector notation to express relationships in linear imaging system and assuming the unknown function \mathbf{f} and the response functions \mathbf{h} belong to Hilbert vector space, \mathbf{g} is identified as an inner product between the vector \mathbf{h} and \mathbf{f} . This may be expressed in vector notation by the following system of linear equations:

$$\mathbf{g}_j = \mathbf{h}_j^T \mathbf{f}, \ j = 1, ..., P \tag{3.1}$$

where, P is the total number of detector tubes and the superscript T denotes the transpose operation, which converts a column vector into a row vector. All equations implied by (3.1) may be summarized by a single matrix equation as:

$$\mathbf{g} = \mathbf{H}\mathbf{f}.\tag{3.2}$$

Here, \mathbf{h}_j is the *j*-th row of the projection matrix \mathbf{H} , and each element of \mathbf{f} , denoted by \mathbf{f}_i , i = 1, ..., N, represents one pixel in the image space. The point-spread function $h_i(\mathbf{x})$ depicts the effects of attenuation and all linear sources of blur.

Each pixel is correlated with a basis function $b_i(\mathbf{x})$, which converts the continuousdomain function into pixel values. The most regular pixel basis functions are those that are constant within small, non overlapping rectangular regions systematized in a rectangular grid [12]. In this case, the pixels are intended to interpret the standard number of emissions from within that pixel. However, there are other appropriate basis functions, such as Gaussian basis functions or finite-element models that adapt to the content of the image.

The projection area is also discrete, with the projection data symbolized by the vector \mathbf{g} . Elements of \mathbf{g} are remarked as projection bins, and every projection measurement is matched with one bin. In addition, bins are generally sampled uniformly.

Since the data are innately discrete and the detection process linear, the mapping between the source image and the expected value of the true coincidence data can be represented by a forward projection matrix, \mathbf{H} , which appears in (3.2) (Figure 3.3). The system matrix \mathbf{H} is a $P \times N$ matrix, with elements, \mathbf{h}_{ji} , containing the probabilities of detecting an emission from pixel site *i* at detector pair *j*. It is in the stipulation of \mathbf{H} that the projection process can become simple or composite because the measurement of a projection bin is a weighted sum of intensities of the image pixels. In the Radon case, the matrix elements are determined according to the concern that a projection bin accepts contributions only from pixels that are intersected by a given ray and the contributions of other pixels are set to zero. A more sensible case is presented by the linear model wherein a projection bin receives contributions from many weighted pixels, in accordance to the affectability of the projection bin to each pixel. These contributions are related to attenuation, detector response, and scatter and can be estimated from knowledge of the system design.



Figure 3.3: Illustration of a single element of the system model **H**.

3.3 Statistical Model of Event Counts

The basic process of iterative reconstruction is to discretize the image into pixels and treat each pixel value as an unknown. Then a system of linear equations can be set up according to the imaging geometry and physics. Nevertheless, we have concerned just the average behavior of the imaging system and have ignored the variability in the photon-counting process. Taking into account the randomness in the projection data, (3.2) should be written as:

$$E[\mathbf{g}] = \mathbf{H}\mathbf{f} \tag{3.3}$$

where $E[\cdot]$ denotes the expectation operator.

3.3.1 The Poisson model

Proper statistical modeling constitutes the base of statistical iterative reconstruction. The statistical models relates how the projection measurements vary around their expected mean values and is derived from our basic understanding of the acquisition process. Statistical iterative reconstruction leads to good bias and variance attainment in nuclear medicine. The Poisson noise model is critical to accurate analysis of photon-limited image data. In this model the numbers of events detected in the projection bins are independent of one another, thereby, the Poisson probability law states that the probability the random vector of Poisson distributed photon counts equals the true photon counts, given a vector of emission rates, \mathbf{f} , is formulated as:

$$p(\mathbf{g}; \mathbf{f}) = \prod_{j=1}^{p} \frac{\overline{\mathbf{g}_{j}}^{\mathbf{g}_{j}} exp(-\overline{\mathbf{g}_{j}})}{\mathbf{g}_{j}!}$$
(3.4)

where $\overline{\mathbf{g}_{j}}$ is the *j*-th element of $E[\mathbf{g}] = \mathbf{H}\mathbf{f}$:

$$\overline{\mathbf{g}_j} = \sum_{i=1}^N \mathbf{h}_{ji} \mathbf{f}_i \tag{3.5}$$

The Poisson model is a good description of tomographic data and is widely used in tomography field. Applications of the linear Poisson image model, are obtained under photon-limited imaging conditions, emission tomography, gamma-ray astronomy, and fluorescence microscopy [13]. Also, there are other probability models that are often used, such as approximations of the Poisson, shifted Poisson and Gaussian models, which have been proposed to improve model accuracy and for practical computation reasons.

3.3.2 The Gaussian model

As noticed, another image formation model, the Gaussian noise model arises from the transformation of Poisson data. Gaussian model is only effective in cases in which the number of photon counts is sufficiently high. The Poisson model in (3.4) can be approximated by the following Gaussian probability density function:

$$p(\mathbf{g}; \mathbf{f}) = k \exp\left[-\frac{1}{2} \sum_{j=1}^{p} \frac{(\mathbf{g}_j - \overline{\mathbf{g}_j})^2}{\overline{\mathbf{g}_j}}\right] = k \exp\left[-\frac{1}{2} (\mathbf{g} - \mathbf{H}\mathbf{f})^T \mathbf{C}^{-1} (\mathbf{g} - \mathbf{H}\mathbf{f})\right]$$
(3.6)

where k is a normalizing constant, and $\mathbf{C} = diag(\overline{\mathbf{g}_1}, ..., \overline{\mathbf{g}_P})$ is the covariance matrix of \mathbf{g} . This conjecture is sensibly accurate when the mean numbers of events $\overline{\mathbf{g}_j}$ are 20 or greater. Otherwise, at low counts, the Poisson distribution becomes asymmetric about its peak, whereas the Gaussian distribution is always symmetric. Negative values of the elements of \mathbf{g} have a probability of zero in the Poisson law, though the Gaussian approximation tolerates negative values. In this manner, Poisson based algorithms often have involved constraints of non negativity, on the contrary, Gaussian based algorithms mandate further constraints to achieve non negativity.

3.4 Image reconstruction criteria

Image reconstruction could be viewed as the problem of finding an image that fulfils the constraints prescribed by the measured data and prior knowledge. This is the foundation for a variety of algorithms, which posit that each projection measurement $\mathbf{g}_j = \mathbf{h}_j^T \mathbf{f}$, j = 1, ..., P defines a hyperplane in which the solution \mathbf{f} must lie. The core of all these methods is that they intend to satisfy the known constraints. However, the problem is complicated when the data are corrupted by noise [12].

The feebleness of these techniques is that there is no formula for integrating an explicit statistical model of the data. Although, these methods attracted considerable interest so soon in the development of tomography, they have been substituted by the maximumlikelihood and Bayesian methods described below.

3.4.1 Maximum-Likelihood criterion

The Maximum-Likelihood (ML) criterion is a statistical estimation criterion. In the ML criterion, it is presumed that the probability $p(\mathbf{g}; \mathbf{f})$ for the vector \mathbf{g} is governed by a set of unknown deterministic parameter, the vector \mathbf{f} , which is the object distribution we aim to reconstruct. In this framework, $p(\mathbf{g}; \mathbf{f})$ is called the likelihood function, which is denoted by $\mathcal{L}(\mathbf{f})$. In the maximum likelihood problem, the goal is to find the image, among all possible images, that is the finest estimate of the true image and, consequently, maximizes \mathcal{L} . Explicitly, ML criterion forms a direction on choosing the reconstructed image $\hat{\mathbf{f}}$ to be the object function \mathbf{f} for which the greatest likelihood $p(\mathbf{g}; \mathbf{f})$ is achieved.

In this sense, the ML criterion courts a solution which is statistically consistent with the observed data, and can be stated simply as follows:

$$\hat{\mathbf{f}} = \arg\max_{f} p(\mathbf{g}; \mathbf{f}) \tag{3.7}$$

namely, find the value of \mathbf{f} for which $p(\mathbf{g}; \mathbf{f})$ is greatest. Maximum likelihood estimation is a totally analytic maximization procedure. Moreover, ML estimators and likelihood functions generally have very desirable large sample properties. Initially, they become unbiased minimum variance estimators as the sample size increases. Secondly, they have approximate normal distributions and approximate sample variances that can be calculated and used to generate confidence bounds, which makes them less susceptible to noise than other unbiased estimators [12]. In other words, likelihood functions can be used to test hypotheses about models and parameters.

Unfortunately, tomographic reconstructed images using the ML criterion have two important drawbacks. Despite of the fact that they obtain the least variance among unbiased estimators, the variance is still high, and with small number of failures they can be heavily biased and the large sample optimality properties do not apply. Therefore, a certain amount of bias can be permitted in the reconstructed image by inserting spatial smoothing in the images, which diminishes the noise at the expense of reduced fidelity in the mean. Smoothing can achieved explicitly through Bayesian methods, which are described nextly, or implicitly via stopping rules. Moreover, calculating ML estimators requires specialized software for solving complex non-linear equations. This is overcomed as time passes, as more statistical packages have updated to contain ML estimation analysis capability.

3.4.2 Maximum a Posteriori criterion

The ML method suggest a classical estimation criteria, which rests in the assumption that **f** is unknown but deterministic. Deterministic solutions to the measurement proceed directly from them and not from any other information. They are typically implemented without regard to the impact of noise, which always appears in actual measurements. In contrast, Bayesian approach relies on the presumption that the unknown image **f** to be reconstructed is a random selection from an identifiable ensemble of similar images and can, thence, be labelled by a PDF, $p(\mathbf{f})$, that is known in advance of data collection. This PDF, called the prior, allows the experimenter to adjust the reconstructed image in order to coincide to his anticipation into the process. By using this prior information, it is expected that a significant estimate of the null space component of the reconstruction will be supplied, thereby decreasing the artifacts. The Bayesian methods enable prior information about the structure of the reconstructed object to be employed to estimate the null space components of the solution [14].

Bayesian estimation, by incorporating prior information through the choice of a prior distribution for a random field, has already been proved to be an effective solution to the ill-posed reconstruction problem. Based on Bayesian theory, a common contextual constraint can be reshaped into prior knowledge to regularize the solution of the reconstruction problem. Wherefore, the regularization through prior information can be inflicted on image reconstruction process to confine noise efficiently. The ambition of Bayesian analysis is ordinarily to assist the reconstructed image to be smooth, so as to limit the noise effect. Specifically, a low probability is allocate to solutions that have subtle details, which are probably emerged from noise. This surmise is based on the knowledge that, due to blurring effect, the imaging system **H** moderates image features.

Bayes' theory expresses the outcomes of accepting one image solution over another, in the form of a quantity called the loss function, which is denoted by $\lambda(\mathbf{f}, \hat{\mathbf{f}})$. This loss function appreciates the degree to which the reconstructed image $\hat{\mathbf{f}}$ and the true image \mathbf{f} deviate. The loss function utilized in tomography image reconstruction is:

$$\lambda(\mathbf{f}, \hat{\mathbf{f}}) = \begin{cases} 0, & \left| \mathbf{f} - \hat{\mathbf{f}} \right| < \delta \\ 1, & otherwise \end{cases}$$
(3.8)

where δ is a positive constant, and $|\cdot|$ indicates the L_1 norm. This loss function declares that the reconstructed image $\hat{\mathbf{f}}$ is adequate when it is sufficiently close to the true image \mathbf{f} and not admissible when results are less infallible.

Bayesian methods attempt to find the criterion that will minimize the average loss when this criterion is exploited. Minimization of (3.8) leads to the maximum a posteriori criterion, which outlines the pick of the value of **f** that maximizes the posterior PDF, $p(\mathbf{f}, \mathbf{g})$. The MAP estimate is given by:

$$\hat{\mathbf{f}} = \arg\max_{f} p(\mathbf{f}; \mathbf{g}). \tag{3.9}$$

According to Bayes' formula, given by:

$$p(\mathbf{f};\mathbf{g}) = \frac{p(\mathbf{g};\mathbf{f})p(\mathbf{f})}{p(\mathbf{g})},$$
(3.10)

Thence, (3.9) can be rewritten as follows:

$$\hat{\mathbf{f}} = \arg\max_{f} \frac{p(\mathbf{g}; \mathbf{f}) p(\mathbf{f})}{p(\mathbf{g})}.$$
(3.11)

By taking the logarithm of the quantity to be maximized, and ignoring $p(\mathbf{g})$ because it does not depend on \mathbf{f} , the MAP criterion can be simplified to the following form:

$$\hat{\mathbf{f}} = \arg\max_{\mathbf{c}} [\ln p(\mathbf{g}; \mathbf{f}) + \ln p(\mathbf{f})]$$
(3.12)

From (3.12), it seems that maximum likelihood estimation is equivalent to MAP estimation with a prior distribution, which is uniform over the feasible reconstruction set. MAP image reconstruction offers an improvement over ML method, because, it uses the logarithm of the prior to penalize solutions that do not agree with the projected properties. Specifically, the maximization in (3.12) attempts to generate an image that is consistent with the data while not being too noisy.

Generally, the MAP criterion allows us to inject into the estimation our prior beliefs. MAP estimation outweighs ML criterion for an extra reason, which is the sharpness of the function that must be optimized, in opposition with the likelihood function and leads to more efficient iterative reconstruction algorithms. Figure 3.4 shows the effect of the prior on a ML objective function. The sharpness of the prior around its peak determines that the resulting MAP solution approaches the peak of the prior. Obviously, if the peaking of the prior is extreme, then the resulting solution will depend primarily on the prior and the measured data will be ignored which can be avoided by using a weak prior, so that the solution captures the main properties of the ML solution while pushing the solution slightly to a direction that emphasizes smoothness [12].



Figure 3.4: Comparison of objective functions in a simple 1D example [12].

3.5 Iterative Reconstruction Algorithms

Tomographic images have been reconstructed from raw data using filtered back projection since the inception of the modality. The standard FBP algorithm operates on several fundamental conjectures about scanner geometry but is basically a compromise between reconstruction speed and image noise. Various assumptions about scanner geometry are compounded with multiple iterations of reconstruction which is termed iterative reconstruction. Iterative methods improve the analytical approach, because they take into account noise in the observations and can apply a more realistic model. These improvements occur at the expense of complexity, since they yield on mathematical problems without a direct analytic solution or with an analytic solution that cannot be solved with current processing capabilities. Therefore, these more realistic approaches are often replaced by methods that successively improve an estimate of the unknown image. This iterative process results in a potentially more accurate estimate than analytical reconstruction methods, via considerable computational requirements. Progress in computation speed and optimized algorithms have reduced the computational burden of these methods, permitting their growing clinical acceptance. The advantage of the discrete approach for iterative algorithms is that the entire acquisition process, including the interaction of the photons with the object, the collimator and the detector, can be incorporated directly. This is very complicated and therefore different approximations have been proposed to reduce the calculation time [15].

An additional benefit of iterative reconstruction methods is the substantially less image noise from the same raw data through more complex modeling of detector response and of the statistical behavior of measurements. Iterative algorithms ground in an precise system model which incorporates the collimator response function, the attenuation, and the scatter. Iterative image reconstruction methods comprise a criterion for choosing the best image, combined with an algorithm for estimating that image. Many different iterative approaches for solving the tomographic reconstruction problem have been investigated, but they share some common properties. The principle of the iterative algorithms is to compute a solution by iterative estimates. The projections corresponding to the current estimate are compared with the measured projections. The result of the comparison is used to modify the current estimate, thereby creating a new estimate. The algorithms differ in the way the measured and estimated projections are compared and the kind of correction applied to the current estimate.

More precisely, the process initiates with an initial estimate $\hat{\mathbf{f}}^{(0)}$ of the pixel intensities in the image. A projection step is imposed on the current image estimate $\hat{\mathbf{f}}^{(t)}$, which outputs a set of projection values $\hat{\mathbf{g}}^{(t)}$ that would be expected if $\hat{\mathbf{f}}^{(t)}$ were the true image. The resulting projections $\hat{\mathbf{g}}^{(t)}$ are, then, compared with the actual measured data \mathbf{g} in order to create a set of projection-space error values \mathbf{e}_g . These are mapped back to the image space through a back projection process to generate image-space error values \mathbf{e}_f that are used to update the image estimate, which becomes the new estimate $\hat{\mathbf{f}}^{(t+1)}$. Iterative reconstruction algorithms follow the general model shown in Figure 3.5 [12].

This process is repeated until the estimated image does not change significantly or when a predefined number of iterations is reached. At the conclusion of the process, the current image estimate is considered to be the final solution.

A crucial issue in image reconstruction is the practical computational options tangled with the implementation of the forward and back projection steps involved in all iterative



Figure 3.5: Flowchart of a generic iterative reconstruction algorithm [12].

algorithms. In many cases, this topic can preclude the use of certain algorithm types due to their impossible calculation.

It is critical in the projection, comparison, back projection, and update steps that individual reconstruction algorithms differ. The power of iterative methods lies in the use of this feedback loop to refine the reconstructed image. In the following sections, we describe the most predominant iterative algorithms, which are based on statistical estimation criteria.

3.5.1 The Maximum-Likelihood Expectation-Maximization Algorithm

Expectation Maximization (EM) algorithm is a general iterative method. In 1982, Shepp and Vardi proposed an implementation of the EM algorithm adjusted to the problem of image reconstruction in PET. However, the resulting iterative formula had been derived by a different approach in the 1970's which was already known as the Richardson-Lucy algorithm [8], [9]. This is called maximum likelihood expectation maximization and was initially employed as the solution to incomplete data problems in statistics finding application in a wide range of statistical applications. The MLEM algorithm and its variations is proper for emission tomography owing to its accurate mathematical model which perceives the physics of emission tomography from that of transmission tomography.

The name of the MLEM algorithm originates from the expectation step that uses current parameter estimates in order to execute a reconstruction of the unobservable Poisson process. The first step is followed by a maximum likelihood that uses this reconstruction to update the parameter estimates. The maximum likelihood approach in image reconstruction for emission tomography was first introduced by Rockmore and Macovski. Explicitly, MLEM algorithm computes an unknown density distribution in the source from the measured counts. Consequently, it is possible to calculate the probability that any initial distribution density of the object could have yielded to the measured data. The image having the highest probability is the maximum likelihood estimates of the original object.

The MLEM requires multiple iterations depending on the scanner geometry and the measured data. For this purpose, Hudson and Larkin proposed an accelerated version of the original MLEM algorithm based on the grouping of the acquired data in subsets, which is called ordered subsets expectation maximization.

Now let us describe the iterative formula of MLEM in tomographic image reconstruction. The complete data \mathbf{s}_{ji} , which is the number of photons emitted from within pixel i and detected in projection bin j, can be related to the observed projection data \mathbf{g} and the image \mathbf{f} as follows:

$$\mathbf{g}_j = \sum_i \mathbf{s}_{ji},\tag{3.13}$$

$$E[\mathbf{s}_{ji}] = \mathbf{h}_{ji}\mathbf{f}_i. \tag{3.14}$$

The E-step of the EM algorithm requires the complete data log-likelihood, $\ln p(\mathbf{s}; \mathbf{f})$. In ET, the counts \mathbf{s}_{ii} are independent Poisson-distributed random variables, therefore,

$$p(\mathbf{s}; \mathbf{f}) = \prod_{j} \prod_{i} \frac{E[\mathbf{s}_{ji}]^{\mathbf{s}_{ji}} e^{-E[\mathbf{s}_{ji}]}}{\mathbf{s}_{ji}!}$$
(3.15)

and the log likelihood is given by:

$$\ln p(\mathbf{s}; \mathbf{f}) = \sum_{j} \sum_{i} [\mathbf{s}_{ji} \ln(\mathbf{h}_{ji} \mathbf{f}_{i}) - \mathbf{h}_{ji} \mathbf{f}_{i} - \ln(\mathbf{s}_{ji}!)].$$
(3.16)

Using (3.16), the E-step is calculated as follows:

$$\mathbf{Q}(\mathbf{f}, \mathbf{\hat{f}}^{(t)}) = E[\ln p(\mathbf{s}; \mathbf{f}) | \mathbf{g}; \mathbf{\hat{f}}^{(t)}] = \sum_{j} \sum_{i} \{ E[\mathbf{s}_{ji} | \mathbf{g}; \mathbf{\hat{f}}^{(t)}] \ln(\mathbf{h}_{ji} \mathbf{f}_{i}) - \mathbf{h}_{ji} \mathbf{f}_{i} - E[\ln(\mathbf{s}_{ji}!)] \}.$$
(3.17)

(...)

The conditional mean of \mathbf{s}_{ji} in (3.17) is given by:

$$E[\mathbf{s}_{ji}|\mathbf{g}; \hat{\mathbf{f}}^{(t)}] = \frac{\mathbf{h}_{ji} \hat{\mathbf{f}}_{i}^{(t)}}{\sum_{k} \mathbf{h}_{jk} \hat{\mathbf{f}}_{k}^{(t)}} \mathbf{g}_{j} \triangleq x_{ji}, \qquad (3.18)$$

which is simply the fraction of the detected counts in projection bin j that are expected to have emanated from pixel i, given that current image estimate $\hat{\mathbf{f}}^{(t)}$ is the source of these counts. Substituting (3.18) into (3.17), the final form of E-step is obtained:

$$\mathbf{Q}(\mathbf{f}, \mathbf{\hat{f}}^{(t)}) = \sum_{j} \sum_{i} \{ x_{ji} \ln(\mathbf{h}_{ji}\mathbf{f}_{i}) - \mathbf{h}_{ji}\mathbf{f}_{i} - E[\ln(\mathbf{s}_{ji}!)] \}.$$
(3.19)
In the M-step, we estimate $\hat{\mathbf{f}}^{(t+1)}$ by maximizing $\mathbf{Q}(\mathbf{f}, \hat{\mathbf{f}}^{(t)})$ with respect to \mathbf{f} . Thence, we set the derivative of $\mathbf{Q}(\mathbf{f}, \hat{\mathbf{f}}^{(t)})$ to zero:

$$\frac{\partial \mathbf{Q}(\mathbf{f}, \hat{\mathbf{f}}^{(t)})}{\partial \mathbf{f}_i} = \mathbf{0}.$$
(3.20)

Solving for $\hat{\mathbf{f}}_{i}^{(t+1)}$, with i = 1, ..., N, the following simple iterative equation is obtained, which is the MLEM iteration for tomographic image reconstruction:

$$\hat{\mathbf{f}}_{i}^{(t+1)} = \frac{\hat{\mathbf{f}}_{i}^{(t)}}{\sum_{j'} \mathbf{h}_{j'i}} \sum_{j} \mathbf{h}_{ji} \frac{\mathbf{g}_{j}}{\sum_{k} \mathbf{h}_{jk} \hat{\mathbf{f}}_{k}^{(t)}}.$$
(3.21)

where $\hat{\mathbf{f}}_{i}^{(t)}$ is the estimated activity in pixel *i* after the *n*-th iteration, \mathbf{h}_{ji} is an element of *H* that represents the probability that a photon emitted from pixel *i* is detected in detector element *j*.

The MLEM algorithm in (3.21) corresponds to the general model of an iterative algorithm (see Fig. 3.6). As it is illustrated, we start with an initial image guess, which is visible in the denominator of equation (3.21). This initial choice is usually just the entire image set to a constant value. The first step projects this image into the projection domain and, next, these projections are compared with the measured projections. This shapes a correction factor for each projection, which is then back projected into image domain to obtain a correction factor for the initial image estimate. The previous quantity is then multiplied by the current image estimate and divided by a weighting term based on the system model. The new image estimate is now re-entered in the algorithm as the next image. The algorithm is repeated while the estimate approaches the maximum likelihood solution. All pixels are updated at the same time in the MLEM algorithm. Its convergence property is consistent. Moreover, MLEM enforces a non negativity constraint. The implementation of the algorithm is simple and eventually leads to a constrained ML solution.

Non-negativity and self-normalization properties of MLEM, are considered the significant advantages of this methodology, as well as its capability to embed in the probability matrix **H** several physical factors, such as attenuation, scatter and accidental coincidence corrections. As an iterative technique, however, MLEM is based on ML criterion, thus, ascribes to noisy reconstructed images. As the EM algorithm iteratively estimates an image, the low frequency components of the image show up within the first few iterations. As the ML estimate is approximated, high frequency definition is solved in the image, effectively adding more variance, which is manifested as noise, to the reconstruction. This variance is often diminished, at the expense of increased bias, by disrupting the algorithm early or by post-smoothing the reconstruction. MLEM algorithm yields satisfying results, if the procedure is stopped prematurely, and the results may gain from applying a post reconstruction low pass filter. Several approaches for determining when to stop the iterations were proposed. An alternative approach is the method of sieves, which includes smoothing within each iteration to restrict the solution.



Figure 3.6: The maximum-likelihood expectation-maximization algorithm in the form of the general iterative model [12].

The image noise increases as the estimate approaches the ML solution. MLEM provokes, consistently, low spatial frequencies to appear first and then gradually higher spatial frequencies as the iterations progress. Thus, early stoppage of MLEM iterations is equivalent to an implicit smoothing of the reconstructed image. Although the log-likelihood appears to stabilize early in the iterative process, the image estimates continue to change. The log-likelihood function is generally not a suitable measure of image quality for the same reason that the ML criterion is not an ideal reconstruction criterion, and it should be mentioned that images having the same log likelihood value can appear very different.

A second drawback of MLEM is the convergence of the algorithm, which is slow. While the convergence rate of MLEM is image dependent, MLEM usually requires approximately 30-50 iterations to achieve a sufficient solution. MLEM requires one forward projection and one bac kprojection at each iteration, that's why the overall processing time is considerably more than the filtered back projection approach, but leads to a potentially more accurate reconstruction results. The required computation time initially impeded the popularity of the MLEM method in clinical use.

3.5.2 Maximum A Posteriori Reconstruction Algorithms

The reconstructed images acquired by the MLEM algorithm tend to become noisy since the number of iterations increases, wherefore, the ML criterion is not the best criterion to obtain subjectively high quality images. On the other hand, the Bayesian methodology attempts to enhance the quality of the reconstructed image by exploiting the knowledge of the image. This information is called a priori and, using Bayes' rule, is often incorporated into a maximum a posteriori objective function. The a priori knowledge inserts in the iterative procedure a prior term that enforces conditions on the image estimate at each iteration leading to certain convergence. This is equivalent to employing a penalty term at each iteration and, therefore, these methods are also named penalized. The prior (penalty) term can forward desired properties in the image such smoothness levels, edges, or even particular structures based on anatomical side informations.

The introduction of a prior information as a constraint that may foster convergence is called regularization. In addition, the prior is usually chosen to penalize the noisy images. The maximization leads to an iterative scheme called the one-step-late (OSL) algorithm, described by Green:

$$\hat{\mathbf{f}}_{i}^{(t+1)} = \frac{\hat{\mathbf{f}}_{i}^{(t)}}{\sum_{j'} \mathbf{h}_{j'i} + \beta \frac{\partial \mathbf{U}(\mathbf{f})}{\partial \mathbf{f}}} \sum_{j} \mathbf{h}_{ji} \frac{\mathbf{g}_{j}}{\sum_{k} \mathbf{h}_{jk} \hat{\mathbf{f}}_{k}^{(t)}}.$$
(3.22)

This algorithm differs from the MLEM algorithm mainly in that there is a prior term in the denominator. This term is the derivative of an energy function \mathbf{U} , which enforces smoothness and β is a constant that modulates the impact of the prior. It is obvious that if the neighboring pixels of pixel *i* have the same value, then the derivative of energy function tends to zero, so, the new equation which arises is alike to the MLEM formula. Furthermore, if the value for pixel *i* is in average higher than its neighbours, then the derivative is positive and if the value for pixel *i* is in average lower than its neighbors, the intensities of the new image is forced to be higher than it would be if the MLEM was applied.

Several proposed algorithms have been based on the MAP formula, such as the MAP Conjugate Gradient (MAP-CG) algorithm. This algorithm also requires approximately 10-15 iterations, but adds some complexity in order to perform a local linear fit of the prior term in the step size calculation and has no non negativity constraint. It is also possible to use an EM algorithm with a Gaussian likelihood and a Gaussian prior or a Gamma distributed prior. An alternative method is the Preconditioned CG (PCG) algorithm which requires line searching to optimize the step size and does not have the convergence problems of the OSL method.

MAP methods moderate the principal issues associated with ML algorithms. First, MAP reconstructions are smoother and the resulted estimates tend to reach a point at which they change very little with further iterations, indicating approximate convergence. Besides, as the value of β is reduced, the degree of smoothing is limited because this places less weight on the prior. As, β increases, smoothing increases and important image traits degrade. Hence, adjusting the weighting parameter is crucial because large values of β result in less contrast and detail and low values of β produce images that are too noisy. Different types of priors generate different smoothing features.

Although MAP reconstruction successfully smooths noise and improves convergence, it also has several disadvantages. A first drawback is that the denominator can take negative values, and the calculated pixel value may become negative, which is not valid because this value corresponds to the number of photons. This problem may be overcomed by keeping the weighting factor of the prior β low. Another matter is the selection of parameters' values, which is critical. Finally, the prior forces the image to be smooth on the edges, and this results in a loss of image features or, in some cases, the creation of blurry features. This proves that the MAP algorithm adds some bias to the ML problem in exchange for reduced noise variance. To overcome this problem, many other priors have been proposed. The basic assumption is that if there is a large difference between pixel values, then it is likely that the pixel is on an edge, and, thence, the smoothing term has to be as close to zero as possible.

3.5.3 Gibbs prior

The idea of transforming an ill-posed reconstruction problem to a well-posed one grounds in inserting extra inspection on which solutions are more suitable than others. This means that the reconstructed image is required to be consistent with additional criteria, that are setted independently from the data. These restrictions can be considered as Tikhonov regularization, penalty functions, or as Bayesian priors. They are all designed to boost the solution towards a predefined affair about the nature of the true image.

Therefore, the choice of the prior is crucial. The resulting image reflects the assumptions made when constructing the prior. Too severe penalties may cause loss of relevant information, that's why the prior should be as generic as possible, as long as the ill-posedness can be dealt with it.

A smooth image is an image with neighboring pixels having similar intensity values. A prior that encourages this attribute tries to restrict sharp alternations between pixels, because such characteristics appear probably due to noise. A simple mathematical model having this property is the Markov random field, which can be described by the Gibbs pdf:

$$p(\mathbf{f}) = \frac{1}{Z} exp[-\beta \mathbf{U}(\mathbf{f})]$$
(3.23)

with Z being a normalizing constant, called the partition function. The parameter β is a scalar weighting factor that stipulates the spiking of the distribution to its maximum, and $\mathbf{U}(\mathbf{f})$ is the energy function. The non-negative energy function \mathbf{U} has its minimum and the prior has its maximum when the image meets the prior assumptions. A common choice for \mathbf{U} in (3.23) is an energy function with a potential function V involving the differences between pixels in the neighborhood (clique) $S_c(c = 1, ..., C)$:

$$\mathbf{U}(\mathbf{f}) = \sum_{c=1}^{C} \sum_{s_1, s_2 \dots \in S_c} w_{v_1 v_2 v_3 \dots} V_c(\mathbf{f}_{v_1}, \mathbf{f}_{v_2}, \mathbf{f}_{v_3}, \dots)$$
(3.24)

In (3.24), pixels indexed by $v_1, v_2, v_3, ...$, are elements of the same clique. This general model includes various priors proposed for tomographic reconstruction model, including

Gaussian and entropy priors. Cliques may have any number of pixels. In most tomographic reconstruction applications, there is a clique for each pixel i, and each clique consists of pixels that are nearest to pixel i. Methods using two pixel cliques generally use potential functions that are related to the difference in intensity between the two pixels, and the potential functions do not vary across the image. In this manner, (3.24) is simplified to:

$$\mathbf{U}(\mathbf{f}) = \sum_{i=1}^{N} \sum_{j \in S_i, i < j} w_{ij} V_{ij} (\mathbf{f}_i - \mathbf{f}_j)$$
(3.25)

where w_{ij} is the weight of pixel *i* in the neighborhood of pixel *j*.

Gibbs distribution provides a mathematically powerful tool to model a class of priors that specifies the local spatial correlations of the underlying source. Generally, different choices for the potential function may lead to different priors. Although there are many options for the clique structure and weights, usually cliques composed of local neighborhoods and weights are defined by the inverse of the distance between the two pixels in the clique. The principal difference between methods usually lies in the choice of potential function, which determines the smoothness properties of the MAP solution.

Using the Gibbs prior, the log posterior pdf (3.12) becomes:

$$\ln p(\mathbf{f}, \mathbf{g}) = \ln L(\mathbf{f}) - \beta \mathbf{U}(\mathbf{f})$$
(3.26)

where the likelihood function $L(\mathbf{f})$ may have either the Poisson or Gaussian form described earlier. Large values of the likelihood function may be produced by noisy images, but they will be penalized by the prior term $\beta \mathbf{U}(\mathbf{f})$ and thus will not be chosen when maximizing $\ln p(\mathbf{f}, \mathbf{g})$. By these means, there is a balance between the requirements of the measured data and the requirements of the prior through the weighting parameter β . If β is set to zero, the MAP solution is converted to the ML and as β becomes large, the prior term dominate the maximization. The hyperparameter of the optimization, β , is the most significant term for the degree of smoothness in the solution. Usually, β is determined by the user, but there are several methods for its automatic determination.

3.5.4 Total Variation prior

Here, we introduce an other a priori constraint, which is called Total Variation (TV) prior. Using the TV prior, the log posterior pdf (3.12) becomes:

$$\ln p(\mathbf{f}, \mathbf{g}) = \ln L(\mathbf{f}) - \beta \|\mathbf{U}\|_{TV}$$
(3.27)

where, the prior is computed by:

$$\|\mathbf{U}\|_{TV} = \sum_{k=1}^{\sqrt{m}-1} \sum_{l=1}^{\sqrt{m}} |\mathbf{f}_{k,l} - \mathbf{f}_{k+1,l}| + \sum_{k=1}^{\sqrt{m}} \sum_{l=1}^{\sqrt{m}-1} |\mathbf{f}_{k,l} - \mathbf{f}_{k,l+1}|, \qquad (3.28)$$

with $\beta > 0$ being the regularization parameter, and **f** imposing on the constraint $\mathbf{f} \geq 0$, which means that the components of **f** are non negative. It is assumed that **f** is a vector corresponding to a square $\sqrt{m} \times \sqrt{m}$ image for simplicity of presentation. The regularization parameter β is specified by the user. However in practice this parameter can be chosen via a cross-validation procedure.

Regularization based on Total Variation (TV) has gained significant attention. This prior measures how much an image varies across pixels, so that a highly textured or noisy image will have a large TV prior, while a smooth or piecewise constant image would have a relatively small, but the method involves a matrix inverse operation which can be extremely difficult to compute for large problems outside of deconvolution settings [17].

3.5.5 Preconditioned Conjugate Gradient

The EM algorithm alternates between estimating the unobserved variables given the current model and the model given the estimated data. In spite of EM method's tremendous success, due to its simplicity and fast initial progress, its speed convergence tends to be extremely slow. Since the data sizes in image reconstruction problem are extensive, iterative image reconstruction algorithms must converge promptly in order to be appropriate. Many methods have been proposed to improve the convergence speed of EM, mostly based on contractual optimization strategy. All of these approaches, although prosperous in terms of convergence, are more complex than EM, that's why they have not accept much popularity in practice. However, Conjugate-Gradient (CG) algorithms are attractive due to their convergence rate, simplicity, and potential for parallelization.

Gradient-based iterative methods often converge slowly for tomographic image reconstruction problems, but can be accelerated by suitable preconditioners. The aim of preconditioning is to produce a coordinate transformation that improves the condition number of a problem, which leads to faster convergence. The preconditioned CG algorithm has been demonstrated effective in tomography. Several generic preconditioners for CG methods are depicted in textbooks. The most common and perhaps most prevalent preconditioners are diagonal scaling matrices.

Gradient-based optimization methods use the gradient of the log-likelihood of the posterior pdf $\ln p(\mathbf{f}; \mathbf{g})$, to define a series of direction vectors d along which $\ln p(\mathbf{f}; \mathbf{g})$ is minimized via 1D line search. Conjugate-gradient methods alter the search directions to confirm that they are mutually conjugate. The following is the preconditioned form of the Polak-Ribiere CG method [16]:

$$\vec{\chi_n} = -\nabla \ln p(\mathbf{f}_n; \mathbf{g}) \quad (\text{-gradient})$$

$$\psi_n = \mathbf{M}\chi_n \quad (\text{preconditioner})$$

$$\kappa_n = \begin{cases} 0, & n = 0\\ \frac{\langle \chi_n - \chi_{n-1}, \psi_n \rangle}{\langle \chi_{n-1}, \psi_{n-1} \rangle}, & n > 0 \end{cases}$$

$$d_n = \psi_n + \kappa_n d_{n-1} \quad (\text{search direction})$$

$$a_n = \arg \min_a \{ \ln p(\mathbf{f}_n + ad_n; \mathbf{g}) \} \quad (\text{step size})$$

$$\mathbf{f}_{n+1} = \mathbf{f}_n + a_n d_n \quad (\text{update})$$

The matrix \mathbf{M} above is the preconditioner; choosing this matrix is part of the algorithm requirements. In order to converge, the preconditioner must be symmetric positive definite. The classical diagonal preconditioner is simply the inverse of the diagonal elements of Hessian of the objective function, \mathbf{H} :

$$\mathbf{M}_{D}(x) \triangleq \mathcal{D}\left[\frac{1}{\mathbf{H}_{jj}(x)}\right]$$
(3.29)

These diagonal elements are positive since \mathbf{H} is positive definite.

Thus, a PCG algorithm is applied for maximum likelihood estimation in proposed models, and it can surpass EM convergence issues. The main idea of the PCG algorithm is that knowing the posterior we can calculate the exact gradient of the objective function and, hence, this gradient can be used with the preconditioner in any standard manner [16].

Additionally, PCG computes the true gradient and shifts in a search direction determined by the conjugate gradient algorithm, which may be different than the EM step direction. The choice of initial conditions is very important for PCG so as the EM algorithm. Since EM is based on optimizing a lower bound on the likelihood, once EM is encircled in a poor basin of attraction, it can never find a better local optimum. It turns up that direct optimization methods such as PCG may avoid this problem because of the non local nature of the line search.

Chapter 4

Tomographic image reconstruction with spatially varying mixture models

4.1 Introduction

4.2 Spatially Varying Gaussian Mixture Models

- 4.3 Spatially Varying Gamma Mixture Models
- 4.4 Experimental Results

4.1 Introduction

A regular method for overcoming the problem of slow convergence rate and noise propagation is the Bayesian maximum a posteriori or penalized maximum likelihood tomographic reconstruction methods which impose a priori information as a regularization constraint and amplify the spatial smoothness of reconstructed image [18] - [20]. A common model for the prior is the Markov random field (MRF) expressed by the Gibbs distribution and many methods were proposed in that framework, differing on the choice of the potential function [20] - [22]. Particularly, the computational process of the popular Bayesian Gibbs prior relies on the local differences between neighboring pixels. Despite the high image quality and the enhanced convergence speed, reconstructed images are likely to be globally smooth. On the other hand, Gaussian mixture models have been evolved to ameliorate these shortcomings, but they are sensitive to outliers and may lead to excessive sensitivity to few data points.

More recently, in [23], the notion of clustered intensity histogram is introduced in a penalized likelihood method. A monotonically decreasing surrogate objective function resulting in a closed form expression is proposed in [24] while the median root prior was

also used to impose spatial smoothness and stabilize the solution [25]. Finally, a non local prior was designed [26] where the definition of a pixel's neighborhood is broadened.

Herein, we propose four models, as called spatially varying mixture models, which are based on maximum a posteriori estimate and use two different priors: a Gaussian mixture prior and a Gamma mixture prior. Moreover, in order to account for the modeling of edges between image segments, appropriate MRF smoothness priors on the contextual mixing proportions, namely the probabilities of the pixel labels, that take the form of a *line process* are chosen, which model the existence of a boundary by a binary variable that is accordingly switched on and off. Specifically, the first two spatially varying mixture models are based on a MAP tomographic reconstruction formula, which uses a Gaussian mixture prior and considers edge preservation by imposing both a Student's t-distribution (continuous model) and a Bernoulli prior (binary model) of the contextual mixing proportions. We have applied the models proposed in [27] because these types of prior are employed to ensure the preservation of region boundaries. In the related literature, many spatially varying Gaussian mixtures have been proposed and could be applied as priors in tomography [27] - [30].

Subsequently, we present two alternative spatially varying mixture models, which employ a Gamma mixture prior and, also, impose the binary and continuous priors on the local differences of the contextual mixing proportions. In tomography the spatially varying Gamma mixture model with Gamma prior is applicable when the intensity histogram of the object pixels comprises a few peaks. Additionally, this kind of prior pdf leads to certain computational conveniences, including the easy imposition of object positivity.

The proposed methods succeed in capturing spatial coherence and in preserving image boundaries. A significant benefit of the proposed algorithms is the automatic estimation of parameters from the data as they do not require an empirical selection.

The overall algorithm consists of an alternating optimization scheme. One step consists in estimating the parameters of the spatially varying mixture models using the EM algorithm [27] with the image \mathbf{f} being fixed. Having the parameters fixed from the first step, the second step consists in estimating \mathbf{f} by the MAP-EM algorithm update. Algorithm 1 summarizes the different steps. The algorithm stops when the estimated image does not change significantly or when a predefined number of iterations is reached, which is common in this type of alternating optimization methods [23, 25, 26].

The image reconstruction process integrates these four models in a standard MAP-EM iterative algorithm and in a preconditioned conjugate gradient optimizer, where the models' parameters and the unknown image are estimated. Numerical experiments using photon-limited images reveal the supremacy of the methods over standard and state-ofthe-art algorithms

- 1: input: A sinogram \mathbf{g} , a threshold $\boldsymbol{\epsilon}$, MAXiterations.
- 2: output: The unknown image **f**.
- 3: Initialize **f** by an image with constant intensity.
- 4: counter=0.
- 5: while $||\mathbf{f}^{(t+1)} \mathbf{f}^{(t)}|| > \epsilon$ and counter \leq MAXiterations do
- 6: Estimate the parameters of the spatially varying mixture model using the EM algorithm.
- 7: Estimate the image \mathbf{f} using the MAP-EM update in (3.22).
- 8: counter++

Algorithm 1: MAP-EM tomographic reconstruction using a spatially varying mixture model

4.2 Spatially Varying Gaussian Mixture Models

Let **f** be the vector of features (e.g., intensity, textural features, location, etc.) representing an image spatial location (pixel). Viewing the required reconstruction as a clustering problem on **f**, we can assume that the \mathbf{f}_n are independent, identically distributed and that they are generated by a finite mixture model [31]:

$$p(\mathbf{f}_n) = \sum_{j=1}^{K} \pi_j \phi(\mathbf{f}_n; \theta_j)$$
(4.1)

where $\Pi = {\{\pi_j\}_{j=1}^K}$ are the prior probability of a pixel membership on class j, which are called the contextual mixing proportions and are confined to be positive and summing to unity. The ${\{\theta_j\}_{j=1}^K}$ is a set of deterministic parameters controlling the shape of the kernel functions ϕ . Thus, there is a natural correspondence between pixel class-membership and kernels, and we can classify the pixels according to posterior class memberships [27]. A standard and well known choice of kernel function is the Gaussian distribution that is presented in this section with other choices for example the Gamma distribution that is also described later.

We assume that, the conditional distribution of \mathbf{f} , given an explicit hidden variable Z, which leads to significant simplifications for the model, is:

$$p(F|Z) = \prod_{j=1}^{K} \prod_{n=1}^{N} \phi(\mathbf{f}_n; \theta_j)^{z_j^n}$$
(4.2)

while the prior distribution for the latent variable is distributed multinomially [32]:

$$p(Z|\Pi) = \prod_{j=1}^{K} \prod_{n=1}^{N} (\pi_j^n)^{z_j^n}$$
(4.3)

with z^n being a binary vector having a single component equal to 1, $z_j^n = 1$, and all others equal to 0.

The proposed models use a prior density distribution, based on the Gibbs distribution for the random variables Π , to model the intuitive fact that neighboring pixels should be characterized by similar components:

$$p(\Pi) \propto \prod_{c} e^{-\psi_{c}^{(\Pi)}} \tag{4.4}$$

with ψ_c denoting a function on clique c which is known as clique function of the pixel vectors within the clique.

A proper clique distribution choice for local differences of contextual mixing proportions is a Gaussian distribution:

$$\pi_j^n - \pi_j^k \sim \mathcal{N}(0, \beta_{jd}^2), \quad \forall n, j, d, \forall k \in \gamma_d(n)$$
(4.5)

and the joint distribution on Π is given by:

$$p(\Pi;\beta) = \prod_{d=1}^{D} \prod_{j=1}^{K} \prod_{n=1}^{N} \prod_{k \in \gamma_d(n)} \mathcal{N}(\pi_j^n - \pi_j^k, \beta_{jd}^2)$$
(4.6)

with $\mathcal{N}(.)$ being a Gaussian distribution, β_{jd}^2 controls the amount of regularization of the contextual mixing proportions, D defines the number of a pixel's neighborhood adjacency types, K is the number of components and $\gamma_d(n)$ is the neighborhood of pixel n with respect to the d-th adjacency type. In proposed models, 4 neighbors for each pixel (first-order neighborhood) is assumed, and partition the corresponding adjacency types into horizontal and vertical, thus, setting D = 2 (see Fig.4.1 for a detailed illustration). This variability of parameter sets aims to capture the fact that smoothness statistics may vary along clusters and along different spatial directions. In other words, parameters β_{jd}^2 can be used to express not only the class variance for clusters but also the variance within clusters at a certain spatial direction D.



Figure 4.1: First-order neighborhood cliques in the contextual mixing proportions mesh. (a) Each MRF site is associated with a probability scalar value π_j^n and is dependent on 4 neighbors. (b) The set of horizontal neighbors, $\gamma_1(n)$, is highlighted. (c) The set of vertical neighbors, $\gamma_2(n)$, is highlighted [32].

In the current work two smoothing priors for the local contextual mixing proportion differences are used and described in the following Sections. The local differences are assumed to be associated with a set of hidden random variables u, called in the literature line process.

4.2.1 Continuous, Gamma Distributed Line Process Model

The K-kernel Spatially Varying GMM, in contrast with the standard GMM is characterized by the dependence of its mixing proportions, generally called contextual mixing proportions. Specifically, each pixel \mathbf{f}_n , n = 1, ..., N has a distinct mixing proportions vector π_j^n , j = 1, ..., K which denotes the probability of the *n*-th pixel to belong to the *j*-th component [32]. Assuming that the independently distributed random variables \mathbf{f} are a realization of a Gaussian mixture model (GMM), the probability of a single pixel is expressed by:

$$p(\mathbf{f}_n; \pi, \mu, \Sigma) = \sum_{j=1}^J \pi_j^n \mathcal{N}(\mathbf{f}_n; \mu_j, \Sigma_j)$$
(4.7)

where Π is constrained to be positive, $0 \le \pi_j^n \le 1$, and summing to unity, $\sum_{j=1}^J \pi_j^n = 1$ for j = 1, 2, ..., J and n = 1, 2, ..., N. Furthermore, $\mathcal{N}(\cdot)$ is a Gaussian distribution with μ_j the Gaussian kernel mean vector and Σ_j the Gaussian kernel covariance matrices [30]. Hence, the probability of the image is computed by assuming pixel independence, which is common in modeling images by mixtures of distributions:

$$p(\mathbf{f}) = \prod_{n=1}^{N} \sum_{j=1}^{J} \pi_{j}^{n} \mathcal{N}(\mathbf{f}_{n}; \mu_{j}, \Sigma_{j})$$
(4.8)

Apart from enforcing pixel clustering, this prior preserves the edges in the image because the local differences of the contextual mixing proportions are considered to follow a univariate Student's **t**-distribution. Following the definition of the Student's **t**-distribution [33], a two step generative model provides the clique potential functions:

$$\pi_j^n - \pi_j^k \sim \mathcal{N}(0, \beta_{jd}^2 / u_j^{nk}), \qquad (4.9)$$

$$u_j^{nk} \sim \mathcal{G}(\nu_{jd}/2, \nu_{jd}/2), \ \forall n, j, d, \ k \in \gamma_d(n),$$
(4.10)

where $\mathcal{G}(\cdot)$ is the Gamma distribution, $\gamma_d(n)$ is the set of neighbors of the pixel indexed by n, with respect to the d^{th} adjacency type (e.g. horizontal, vertical, diagonal). This model first draws u_j^{nk} from a Gamma distribution parameterized by ν_{jd} and then considers that the local differences of the mixing proportions follow a Gaussian distribution with zero mean and standard deviation β_{jd}^2/u_j^{nk} .

Hence, the student's **t**-distribution's probability density function can be written in the form:

$$p(\mathbf{f};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\nu}) = \frac{\Gamma\left(\frac{\boldsymbol{\nu}+d}{2}\right)|\boldsymbol{\Sigma}|^{-\frac{1}{2}}}{(\pi\boldsymbol{\nu})^{\frac{d}{2}}\Gamma\left(\frac{\boldsymbol{\nu}}{2}\right)\left[1+\boldsymbol{\nu}^{-1}\delta(\mathbf{f},\boldsymbol{\mu};\boldsymbol{\Sigma})\right]^{\frac{\boldsymbol{\nu}+d}{2}}}$$
(4.11)

with $\delta(\mathbf{f}, \boldsymbol{\mu}; \boldsymbol{\Sigma})$ being the Mahalanobis squared distance and Γ being the Gamma function. The **t**-distribution is symmetric well-shaped, and as we observe in Figure 4.2, has heavy tails which means that it is more likely to generate values that fall far from its mean. Additionally, as the number of degrees of freedom rises, the t-distribution approaches the normal distribution with covariance $\boldsymbol{\Sigma}$. If $\boldsymbol{\nu} > 1$, $\boldsymbol{\mu}$ is the mean of \mathbf{f} and if $\boldsymbol{\nu} > 2$, $\boldsymbol{\nu}(\boldsymbol{\nu}-2)\boldsymbol{\Sigma}$ is the covariance matrix of \mathbf{f} .



Figure 4.2: The Student's **t**-distribution for various degrees of freedom.

This generative model, whose graphical representation is shown in Fig. 4.3, allows clustering of the image pixels around the Gaussian means and imposes edge preservation through the Student's **t**-distribution of the mixing proportions. More specifically, as $u_j^{nk} \to +\infty$ the distribution tightens around zero, and enforces neighboring contextual mixing proportions to be smooth. On the other hand, when $u_j^{nk} \to 0$ the distribution tends to be uninformative, and enforces no smoothness. Consequently, the variables u_j^{nk} provide a very detailed description of the boundary structure of the image. Estimation of model parameters through a standard MAP-EM approach is intractable due the complexity of the model and the suitable framework is provided by variational inference which yields estimates for all of the parameters of the model [27].

To perform our formula inference we exploit MAP estimation through EM algorithm. Thus, the E-step of the EM algorithm requires the calculation of the hidden variables Zand u separately with respect to current iteration t parameters :

$$\langle z_{j}^{n} \rangle^{(t)} = \frac{\pi_{j}^{n(t)} \mathcal{N}(\mathbf{f}_{n}; \mu_{j}^{(t)}, \Sigma_{j}^{(t)})}{\sum_{l=1}^{K} \pi_{l}^{n(t)} \mathcal{N}(\mathbf{f}_{n}; \mu_{l}^{(t)}, \Sigma_{l}^{(t)})},$$
(4.12)

$$\langle u_j^{nk} \rangle^{(t)} = \zeta^{nk(t)} / \eta^{nk(t)},$$
 (4.13)

$$\left\langle \ln u_j^{nk} \right\rangle^{(t)} = \psi(\zeta^{nk(t)}) - \ln \eta^{nk(t)}, \qquad (4.14)$$



Figure 4.3: Graphical model for the edge preserving model. Superscripts and subscripts $n, k \in [1, N]$ denote pixel index, subscript $j \in [1, K]$ denotes segment index, $d \in [1, D]$ describes the neighborhood direction type. Γ equals the maximum number of possible neighbors. The Figure is reproduced by [27]

with $\psi(\cdot)$ corresponding to the digamma function while z, η are determined by:

$$\zeta^{nk(t)} = \frac{1}{2} (\nu_{jd}^{(t)} + 1), \qquad (4.15)$$

$$\eta^{nk(t)} = 2\left(\nu_{jd}^{(t)} + \frac{(\pi_j^{n(t)} - \pi_j^{k(t)})^2}{\beta_{jd}^{2(t)}}\right).$$
(4.16)

Afterwards, the unknown variables μ , Σ , β , ν are considered as parameters and computed in the M-step of the algorithm individually:

$$\mu_j^{(t+1)} = \frac{\sum_{n=1}^N \langle z_j^n \rangle^{(t)} \mathbf{f}_n}{\sum_{n=1}^N \langle z_j^n \rangle^{(t)}},\tag{4.17}$$

$$\Sigma_{j}^{(t+1)} = \frac{\sum_{n=1}^{N} \langle z_{j}^{n} \rangle^{(t)} (\mathbf{f}_{n} - \mu_{j}^{(t+1)}) (\mathbf{f}_{n} - \mu_{j}^{(t+1)})^{T}}{\sum_{n=1}^{N} \langle z_{j}^{n} \rangle^{(t)}}, \qquad (4.18)$$

$$\beta_{jd}^{2(t+1)} = \frac{\sum_{n=1}^{N} \sum_{k \in \gamma_d(n)} \langle u_j^{nk} \rangle^{(t)} (\pi_j^{n(t)} - \pi_j^{k(t)})^2}{\sum_{n=1}^{N} |\gamma_d(n)|},$$
(4.19)

In addition, the degrees of freedom, $\nu_{jd}^{(t+1)}$, are estimated as the solution of the bellow equation:

$$\ln\left(\nu_{jd}^{(t+1)}/2\right) - \psi(\nu_{jd}^{(t+1)}/2) + \frac{\sum_{n=1}^{N} \sum_{k \in \gamma_d(n)} \left(\left\langle \ln u_j^{nk} \right\rangle^{(t)} - \left\langle u_j^{nk} \right\rangle^{(t)}\right)}{\sum_{n=1}^{N} |\gamma_d(n)|} + 1 = 0 \quad (4.20)$$

Thereafter, in the same manner, the unknown random variables Π , are treated as parameters and reckoned as the root of the following quadratic equation:

$$a_j^n (\pi_j^{n(t+1)})^2 + b_j^n (\pi_j^{n(t+1)}) + c_j^n = 0$$
(4.21)

with coefficients:

$$a_{j}^{n} = -\sum_{d=1}^{D} \{\beta_{jd}^{-2t} \sum_{k \in \gamma_{d}(n)} \langle u_{j}^{nk} \rangle^{(t)}\}, \qquad (4.22)$$

$$b_j^n = \sum_{d=1}^D \{\beta_{jd}^{-2t} \sum_{k \in \gamma_d(n)} \langle u_j^{nk} \rangle^{(t)} \pi_j^{k(t)} \},$$
(4.23)

$$c_j^n = \frac{1}{2} \langle z_j^n \rangle^{(t)}. \tag{4.24}$$

The real non negative solutions will not satisfy the notion that they are bound to sum to unity. In order to get proper mixing weight vectors we perform a projection step onto the constraints subspace using the quadratic programming algorithm described in [27].

4.2.2 Binary, Bernoulli Distributed Line Process Model

The clique potential functions, set by (4.5) and (4.6) are determined to be distributed as:

$$\pi_j^n - \pi_j^k | u_j^{nk} = 1 \sim \mathbf{N}(0, \beta_{jd}^2), \quad \forall n, j, d, \forall k \in \gamma_d(n),$$
(4.25)

where a line process set of binary random variables $u = \{u_j^{nk}\}_{k=1..\gamma_d(n)}, n = 1..N, j = 1..K, d = 1..D$ is presumed. This form allocates higher probability on local differences which are close to zero only when there is not an edge between them, otherwise, the respective Gaussian is zeroed and, thus, can not rebound to the total MRF energy. In this manner, differences are tightened only between pixels that are not separated by a boundary. We depict the line process binary variables u_j^{nk} to be iid Bernoulli distributed random variables, ruled by a parameter set $\xi = \xi^1, \xi^2, ..., \xi^{\Gamma}$:

$$p(u|\xi) = \prod_{d=1}^{D} \prod_{n=1}^{N} \prod_{k \in \gamma_d(n)} p(u_j^{nk}|\xi^l) = \prod_{d=1}^{D} \prod_{n=1}^{N} \prod_{k \in \gamma_d(n)} \xi^{lu_j^{nk}} (1-\xi^l)^{(1-u_j^{nk})}$$
(4.26)

with the third product, with respect to k, being $l = \phi(n, k)$. Function $\phi(n, k)$ is defined on site indices n and k, necessarily $k \in \gamma_d(n)$ for some $d \in [1, D]$ or ϕ is undefined, and is equal to an index value in the range $[1, \Gamma]$. For fixed n, ϕ defines a one-toone correspondence between site index k and an index $l \in [1, \Gamma]$ [27]. This means that Bernoulli prior is spatially invariant and depends on the direction to the given neighbor.

In order to make the line process model Bayesian, a Beta distribution, which is the conjugate to the Bernoulli distribution, is imposed on the ξ parameters:

$$p(\xi; \alpha_{\xi 0}, \omega_{\xi 0}) = \prod_{l=1}^{\Gamma} \frac{\Gamma(\alpha_{\xi l 0} + \omega_{\xi l 0})}{\Gamma(\alpha_{\xi l 0})\Gamma(\omega_{\xi l 0})} (\xi^l)^{(\alpha_{\xi l 0} - 1)} (1 - \xi^l)^{(\omega_{\xi l 0} - 1)}, \qquad (4.27)$$

with $\alpha_{\xi 0} = \{\alpha_{\xi l 0}\}_{l=1}^{\Gamma}$, $\omega_{\xi 0} = \{\omega_{\xi l 0}\}_{l=1}^{\Gamma}$. To maintain clique symmetry, we force $\alpha_{\xi l 0}$ having the same value for all l, which correspond to the same adjacency type. For instance,

if $\Gamma = 4$ there are four components in vector ξ having two unique values for the horizontal and the vertical direction.

The graphical model showing the dependencies between variables for this model is presented in Fig.4.4.



Figure 4.4: Graphical model for the binary line process edge preserving model. Superscripts $n, k \in [1, N]$ denotes pixel index, subscript $j \in [1, J]$ denotes kernel (segment) index, $d \in [1, D]$ describes the neighborhood direction type and $l \in [1, \Gamma]$ denotes neighbor index. The Figure is reproduced from [27].

After some manipulation, we obtain the update equations for the model parameters, which are as follows:

$$\langle u_{j}^{nk} \rangle^{(t)} = sig(\ln \mathcal{N}(\pi_{j}^{k(t)}, \pi_{j}^{n(t)}, \beta_{jd}^{2(t)} + \langle \ln \xi^{l} \rangle^{(t)} - \ln \langle (1 - \xi^{l}) \rangle^{t}),$$
 (4.28)

$$\left\langle \ln \xi^l \right\rangle^{(t+1)} = \psi(\alpha^t_{\xi l}) + \psi(\alpha^t_{\xi l} + \omega^t_{\xi l}), \qquad (4.29)$$

$$\left\langle \ln(1-\xi^l)\right\rangle^{(t+1)} = \psi(\omega^t_{\xi l}) + \psi(\alpha^t_{\xi l} + \omega^t_{\xi l}), \qquad (4.30)$$

$$\alpha_{\xi l}^{t} = \alpha_{\xi l0}^{t} + \sum_{j=1}^{K} \sum_{n=1}^{N} \langle u_{j}^{nk} \rangle^{(t)}, \qquad (4.31)$$

$$\omega_{\xi l}^{t} = \omega_{\xi l0}^{t} + \sum_{j=1}^{K} \sum_{n=1}^{N} \left\langle 1 - u_{j}^{nk} \right\rangle^{(t)}, \qquad (4.32)$$

$$\forall n, j, d, \forall k \in \gamma_d(n), l = \phi(n, k)$$
(4.33)

where $\psi(\cdot)$ is the digamma function and $sig(x) = (1 + e^{-x})^{-1}$. All other parameters are equal with those in the previous model.

4.3 Spatially Varying Mixture Models with Gamma Mixture Prior

From now on we make the assumption that our image data are generated by a Gamma mixture model, and subsequently extend this by selecting similar prior distributions on

 π_j , as it was presented in the previous section. The model relies on the same principle which was already used for mixtures of Gaussian. The Gamma mixture prior is cluster and directionally adaptive and all its parameters are automatically estimated from the image.

Also, the random variable (image intensity) of a Gamma distribution is by definition non negative and therefore the positivity constraint is inherently satisfied. The general form of a mixture-of-gammas prior pdf is:

$$p(\mathbf{f}|\theta) = \prod_{n=1}^{N} \sum_{j=1}^{K} \pi_j \mathcal{G}(\mathbf{f}_n | q_j, r_j).$$
(4.34)

In (4.34), the set of parameters θ involves the vector q, with elements $\{q_j\}_{j=1,\dots,K}$ and the vector r, $\{r_j\}_{j=1,\dots,K}$, which parameterize the gamma density. The gamma density is defined only for $\mathbf{f} > 0$. Further, $r_j > 0$ is the mean and r_j^2/q_j the variance of the j-th component. Like other positivity preserving priors, the aforementioned prior motivates a slight bias due to the difference between mean and mode. The mixing proportions (weights) π_j are positive and satisfy the constraint:

$$\sum_{j=1}^{K} \pi_j = 1. \tag{4.35}$$

4.3.1 Continuous, Gamma Distributed Line Process Model

This generative model, whose graphical representation is shown in Fig. 4.5, imposes edge preservation through the Student's **t**-distribution on the difference of the mixing proportions. The variables u_j^{nk} provide a very detailed description of the boundary structure of the image. Estimation of model parameters is obtained through a standard MAP-EM approach.



Figure 4.5: Graphical representation of the proposed model.

The final updates differ from the corresponding continuous model with a Gaussian mixture in the computation of the joint expectation of the hidden variables Z and the

parameters $\theta^{(t)}$:

$$\langle z_{j}^{n} \rangle^{(t)} = \frac{\pi_{j}^{n(t)} \mathcal{G}(\mathbf{f}_{n}; q_{j}^{(t)}, r_{j}^{(t)})}{\sum_{l=1}^{K} \pi_{l}^{n(t)} \mathcal{G}(\mathbf{f}_{n}; q_{l}^{(t)}, r_{l}^{(t)})},$$
(4.36)

$$r_j^{(t)} = \frac{\sum \langle z_j^n \rangle^{(t)} \mathbf{f}_n}{\langle z_i^n \rangle^{(t)}}$$
(4.37)

$$q_{j}^{(t)} = \sum_{j=1}^{K} \log \mathbf{f}_{n} \langle z_{j}^{n} \rangle^{(t)} - \langle z_{j}^{n} \rangle^{(t)} \log r_{j}^{(t)} - \langle z_{j}^{n} \rangle^{(t)} \psi(q_{j}^{(t)}), \qquad (4.38)$$

with $\psi(\cdot)$ being the digamma function. The rest of the update equations are the same.

4.3.2 Binary, Bernoulli Distributed Line Process Model

The graphical illustration of the model in Figure 4.6 indicates the dependencies between variables for the binary line process model.



Figure 4.6: Graphical model for the binary line process edge preserving model.

The update equations for the joint expectation of the hidden variables Z and the parameters θ are the same with those in the spatially varying Gamma mixture model described in Section 4.3.1. The rest updates remain the same as in the spatially varying Gaussian mixture model in Section 4.2.2

4.4 Experimental results

The performance of the proposed spatially varying mixture models for the tomographic reconstruction problem were examined using the well known Shepp-Logan phantom and a phantom consisting of three regions of relative intensities, represented by a hot disk, a cold disk and a background ellipse (Fig. 4.7). We call this phantom Elliptical phantom.

We have set K = 3 and K = 5 clusters for the mixture models taking into account the segments of the two phantoms. The algorithm stopped when $\epsilon = 10^{-3}$ or when 60 iterations were reached. The presented algorithms, namely the continuous spatially varying Gaussian mixture model (GMM-CLP), the binary spatially varying Gaussian mixture model (GMM-DLP), the continuous spatially varying Gamma mixture model (GAMMA-CLP) and the binary spatially varying Gamma mixture model (GAMMA-DLP) were evaluated with respect to the standard MLEM, the established MAP-EM algorithm with a Gibbs [20] and a TV prior. In addition, a standard Gaussian mixture model and a standard Gamma mixture model are performed for comparison with the proposed models. Apart from using the one-step-late EM for the estimation of the image, the optimization was also carried out by an iterative preconditioned conjugate gradient (PCG) algorithm with a diagonal preconditioner, which reaches a local maximum as the mixture model is not convex with respect to **f**.

A number of performance indices were used. To this end, degraded images were generated from the initial images by modifying the total photon counts. More specifically, images having 75, 55, 35 and 15 photons/pixel on average were generated to degrade the signal quality, and for the Elliptical phantom, 80, 56, 36 and 24 photons/pixel were simulated.



Figure 4.7: (a) Shepp - Logan phantom. (b) Elliptical phantom.

At first, the algorithms were put in test in terms of the improvement in signal to noise ratio (ISNR) with respect to a reconstruction obtained by a simple filtered back-projection using the Ram-Lak filter:

ISNR =
$$10 \log_{10} \frac{||\mathbf{f} - \mathbf{f}_{FBP}||^2}{||\mathbf{f} - \hat{\mathbf{f}}||^2}$$
, (4.39)

where \mathbf{f} is the ground truth image, \mathbf{f}_{FBP} is the reconstructed image by filtered backprojection and $\hat{\mathbf{f}}$ is the reconstructed image using the proposed image model. Practically, ISNR measures the improvement (or deterioration) in the quality of the reconstruction of the proposed method with respect to the reconstruction obtained by filtered backprojection. Moreover, the consistency of the method was measured by the bias (BIAS) and the variance (VAR) of the reconstructed images:

$$BIAS = ||\mathbf{f} - \bar{\mathbf{f}}||, \qquad (4.40)$$

$$\operatorname{VAR} = \sum_{k=1}^{M} ||\bar{\mathbf{f}} - \hat{\mathbf{f}}_k||^2, \qquad (4.41)$$

with

$$\bar{\mathbf{f}} = \frac{1}{M} \sum_{k=1}^{M} \hat{\mathbf{f}}_{k,k}$$
(4.42)

where **f** is the ground truth image and $\hat{\mathbf{f}}_k$, for k = 1, ..., M, is the k^{th} reconstructed image, obtained from M = 40 different realizations for each noise level. Finally, we also included in the evaluation the structural similarity index (SSIM) [34], which represents the visual distortion between the ground truth and the reconstructed image:

SSIM(
$$\mathbf{f}, \hat{\mathbf{f}}$$
) = $\frac{(2\mu_{\mathbf{f}}\mu_{\hat{\mathbf{f}}} + C_1)(2\sigma_{\mathbf{f}\hat{\mathbf{f}}} + C_2)}{(\mu_{\mathbf{f}}^2 + \mu_{\hat{\mathbf{f}}}^2 + C_1)(\sigma_{\mathbf{f}}^2 + \sigma_{\hat{\mathbf{f}}}^2 + C_2)}$, (4.43)

where $\mu_{\mathbf{f}}$ and $\mu_{\hat{\mathbf{f}}}$ denote the mean intensity of the ground truth and the estimated image, $\sigma_{\mathbf{f}}$ and $\sigma_{\hat{\mathbf{f}}}$ are the standard deviations of the two images, $\sigma_{\mathbf{f}\hat{\mathbf{f}}}$ is the covariance of \mathbf{f} and $\hat{\mathbf{f}}$ and C_1 and C_2 are constants added to avoid instability. The above statistics are calculated locally on equally sized windows centered at each image pixel and the average values over all pixels are reported here.

The statistical comparisons for the aforementioned algorithms are shown in Figures 4.8, 4.9, 4.10 and 4.11 for the Shepp-Logan phantom. For all the quantities, their mean values over the M = 40 experiments are shown. All of the obtained ISNR and SSIM values are very close to the mean values as their standard deviations over the whole set of experiments are very small. For the ISNR, SSIM, bias and variance for the Shepp-Logan phantom, for 75, 55, 35 and 15 photons per pixel on average are shown.



Figure 4.8: Comparative statistics for various performance indices for the Shepp-Logan phantom for 75 photons per pixel. (a) ISNR (mean values of the 40 experiments), (b) structural similarity (mean value), (c) bias, (d) variance.



Figure 4.9: Comparative statistics for various performance indices for the Shepp-Logan phantom for 55 photons per pixel. (a) ISNR (mean values of the 40 experiments), (b) structural similarity (mean value), (c) bias, (d) variance.



Figure 4.10: Comparative statistics for various performance indices for the Shepp-Logan phantom for 35 photons per pixel. (a) ISNR (mean values of the 40 experiments), (b) structural similarity (mean value), (c) bias, (d) variance.



Figure 4.11: Comparative statistics for various performance indices for the Shepp-Logan phantom for 15 photons per pixel. (a) ISNR (mean values of the 40 experiments), (b) structural similarity (mean value), (c) bias, (d) variance.

The statistical comparisons for the Elliptical phantom for 80, 56, 36 and 24 photons per pixel are illustrated in Figures 4.12, 4.13, 4.14 and 4.15.

It may be observed from all these indices for both phantoms that the proposed spatially varying mixture models reveal a better performance with respect to the other priors. As it can be observed in these figures, as the noise decreases (the number of photon counts per pixel increases) the ISNR becomes larger with the spatially varying Gamma mixture method providing better results. Specifically, for the Shepp-Logan phantom with 75 photons per pixel and for the Elliptical phantom with 80 photons, ISNR reaches its peak through the spatially varying Gamma mixture model with continuous prior carried out by a preconditioned conjugate gradient optimizer (GAMMA-CLP (PCG)), as shown in Figures 4.8(a) and 4.12(a). Moreover, as the noise increases in both phantoms, the spatially varying Gamma mixture methods have similar ISNR values but always larger than the other algorithms.

The same stands for the bias which yields its minimum through the GAMMA-CLP (PCG) for both phantoms as it can be observed in Fig. 4.8(c) and Fig. 4.12(c). Furthermore, the variance of the estimates is relatively consistent for the spatially varying mixture models for both phantoms and for all amounts of photons per pixel, which is due to the clustering effect of the priors. In all these indices, the results illustrate the effectiveness of the proposed models.

Since, the gamma-mixture prior model's pixel values having a clustered histogram, the ability of this model to adapt to the data is confirmed, and at the same time spatial smoothness is yielded. Furthermore, the performance of the TV prior is worth noticing.



Figure 4.12: Comparative statistics for various performance indices for the Elliptical phantom for 80 photons per pixel. (a) ISNR (mean values of the 40 experiments), (b) structural similarity (mean value), (c) bias, (d) variance.



Figure 4.13: Comparative statistics for various performance indices for the Elliptical phantom for 56 photons per pixel. (a) ISNR (mean values of the 40 experiments), (b) structural similarity (mean value), (c) bias, (d) variance.



Figure 4.14: Comparative statistics for various performance indices for the Elliptical phantom for 36 photons per pixel. (a) ISNR (mean values of the 40 experiments), (b) structural similarity (mean value), (c) bias, (d) variance.



Figure 4.15: Comparative statistics for various performance indices for the Elliptical phantom for 24 photons per pixel. (a) ISNR (mean values of the 40 experiments), (b) structural similarity (mean value), (c) bias, (d) variance.

In general, the images provided by the spatially varying mixture models are sharper. Since the GAMMA-CLP (PCG) has the best performance for both phantoms, estimated images obtained by this model, for both phantoms, are illustrated along with the estimates obtained by MLEM and MAP-EM with a Gibbs prior and a TV prior in Figures 4.16 -4.23.



Figure 4.16: The estimated images for the Shepp-Logan phantom with 75 photons per pixel.



Figure 4.17: The estimated images for the Shepp-Logan phantom with 55 photons per pixel.



Figure 4.18: The estimated images for the Shepp-Logan phantom with 35 photons per pixel.



Figure 4.19: The estimated images for the Shepp-Logan phantom with 15 photons per pixel.



Figure 4.20: The estimated images for the Elliptical phantom with 80 photons per pixel.



Figure 4.21: The estimated images for the Elliptical phantom with 56 photons per pixel.



Figure 4.22: The estimated images for the Elliptical phantom with 36 photons per pixel.



Figure 4.23: The estimated images for the Elliptical phantom with 24 photons per pixel.



Figure 4.24: Comparison of horizontal profiles between the original Shepp-Logan phantom and the reconstructed images provided by the proposed GAMMA-CLP (PCG) and the Gibbs prior for 75 counts per pixel.



Figure 4.25: Comparison of horizontal profiles between the original Elliptical phantom and the reconstructed images provided by the proposed GAMMA-CLP (PCG) and the Gibbs prior for 75 counts per pixel.

In addition, to highlight the accuracy of the proposed model, the estimated image intensities along a scan line for both phantoms are shown in Figures 4.24 and 4.25 for the GAMMA-CLP (PCG) model and a MAP-EM with Gibbs prior, where it can be seen that spatially varying Gamma mixture model provides values which are closer to the ground truth. Finally, the execution time of the presented algorithms on a standard PC using MATLAB without any optimization are illustrated in Figure 4.26. It is shown that the proposed models require on average 4 minutes, which is explained by the number of required computations for the EM for the reconstruction and the EM for the estimation of the model parameters. The experimental results are shown in more detail in the Appendix.



Figure 4.26: Execution times for the compared algorithms (60 iterations, $\epsilon = 10^{-3}$).

Chapter 5

Conclusions and future work

The main goal of this work was to explore the possibility of the development of a robust model for tomographic image reconstruction purposes. We have proposed four alternative hierarchical and spatially constrained mixture models in order to enforce them to preserve image edges. Firstly, we have presented a spatially varying Gaussian mixture model with two MRF priors on the contextual mixing proportions: a Student's \mathbf{t} (continuous line process) and a Bernoulli (binary line process) prior. Then, we employed a spatially varying mixture model, which is based on a Gamma mixture prior. Spatially varying mixture models are characterized by the dependence of their mixing proportions on location (contextual mixing proportions) and they have been successfully used in image segmentation. The binary line process affects the model's sensitivity to preserve region boundaries. Moreover, the continuous line process model is computationally and conceptually more simple.

On the other hand, spatially varying Gamma mixture model is proven an apt model due to the gamma mixture prior which assumes a clustered intensity histogram in the object, and enforces positivity naturally.

A property of the proposed models is the automatic estimation of model parameters from the data which is crucial, as many state-of-the-art reconstruction algorithms rely on empirical parameter selection.

Numerical experiments on various photon limited image scenarios showed that the proposed models are more accurate than the widely used Gibbs prior. The main contribution of this work is the effectiveness of the MRF priors which may capture spatial coherence and preserve image boundaries, without imposing smoothness across them.

In this work the ISNR has been used as an indicator of the image quality along the iterative image reconstruction process. This indicator expresses the normalized difference between the phantom and the reconstructed image. The spatially varying Gamma mixture models have the highest ISNR. It has been shown here that this observation is valid independently of the image topology and activity distribution.

This work intends to provide the reader with a starting point for further study, especially, in exploring effective ways of analysing the reconstruction model's convergence property. An important perspective of this study is to automatically estimate the number of components of the mixture model in the image reconstruction framework
Bibliography

- [1] A. C. Kak and M. Slaney, Principles of Computerized Tomographic Imaging, Society of Industrial and Applied Mathematics, 2001.
- [2] A. K. Jain, Fundamentals of Digital Image Processing, Prentice-Hall, 1989.
- [3] J. Radon, On the determination of functions from their integrals along certain manifolds, IEEE Transactions Medical Imaging, vol. 5, no. 4, pp. 170-176, 1986.
- [4] R. C. Gonzalez and R. E. Woods, Digital Image Processing, 3rd Edition, Prentice Hall, 2008.
- [5] S. Webb, The Physics of Medical Imaging, 1rd Edition, 1988.
- [6] R. M. Leahy and Q. Jinyi, Statistical approaches in quantitative positron emission tomography, Journal: Statistics and Computing, 2000.
- [7] G. L. Zeng, Image reconstruction a tutorial, Computerized Medical Imaging and Graphics, vol. 25, no. 2, pp. 97-103, 2001.
- [8] R. W. Hadley, Bayesian-Based Iterative Method of Image Restoration, Journal of the Optical Society of America, vol. 62, no. 1, pp. 55-59, 1972.
- [9] L. B. Lucy, An iterative technique for the rectification of observed distributions, Astronomical Journal, vol. 79, no. 6, pp. 745-754, 1974.
- [10] H. M. Hudson and R. S. Larkin, Accelerated image reconstruction using ordered subsets of projection data, IEEE Transactions on Medical Imaging, vol. 24, no. 7, pp. 797-808, 1994.
- [11] J. A. Browne and A. R. De Pierro, A row-action alternative to the EM algorithm for maximizing likelihood in emission tomography, IEEE Transactions on Medical Imaging, vol. 15, no. 4, pp. 687-699, 1996.
- [12] M. N. Wernick and J. N. Aarsvold, Emission Tomography: The Fundamentals of PET and SPECT, Elsevier Science, USA, UK, 2004.

- [13] T. J. Hebert and R. Leahy, Statistic-based MAP image reconstruction from Poisson data using Gibbs priors., IEEE Transactions on Medical Imaging, vol. 40, no. 9, pp. 2290-2303, 1992.
- [14] K. M. Hanson, Bayesian and related methods in image reconstruction from incomplete data, in Henry Stark, editor, Image Recovery: Theory and Application, pp. 79-125, Academic, Orlando, 1987.
- [15] S. Vandenberghe, Y. D'Asseler and R. van de Walle et al., Iterative reconstruction algorithms in nuclear medicine, Computerized Medical Imaging and Graphics, vol. 25, no. 2, pp. 105-111, 2001.
- [16] J. A. Fessler and S. D. Booth, Conjugate-gradient preconditioning methods for shiftvariant PET image reconstruction, IEEE Transactions on Image Processing, vol. 8, no. 5, pp. 688-99, 1999.
- [17] M. A. T. Figueiredo and J. M. Bioucas-Dias, Restoration of Poissonian images using alternating direction optimization, IEEE Transactions Image Processing, vol. 19, no. 12, pp. 3133-3145, 2010.
- [18] E. Levitan and G. T. Hergman, A maximum a posteriori probability expectation maximization algorithm for image reconstruction in emission tomography, IEEE Transactions on Medical Imaging, vol. 6, no. 3, pp. 185-192, 1987.
- [19] T. Hebert and R. Leahy, A generalized EM algorithm for 3-D Bayesian reconstruction from Poisson data using Gibbs priors, IEEE Transactions on Medical Imaging, vol. 8, no. 2, pp. 194-202, 1989.
- [20] P. J. Green, Bayesian reconstructions from emission tomography data using a modified EM algorithm, IEEE Transactions on Medical Imaging, vol. 9, no. 1, pp. 84-93, 1990.
- [21] D. M. Higdon, J. E. Bowsher and V. E. Johnson, T. G. Turkington, D. R Gilland, and R. J. Jaczszak, Fully Bayesian estimation of Gibbs hyperparameters for emission computed tomography data, IEEE Transactions on Medical Imaging, vol. 16, no. 5, pp. 516-526, 1997.
- [22] E. U. Mumcuoglu, R. Leahy, S. R. Cherry, and Z. Zhou, Fast gradient-based methods for Bayesian reconstruction of transmission and emission PET images, IEEE Transactions on Medical Imaging, vol. 13, no. 4, pp. 687-701, 1994.
- [23] I. T. Hsiao, A. Rangarajan, and G. Gindi, Joint MAP Bayesian tomographic reconstruction with a Gamma mixture prior, IEEE Transactions on Medical Imaging, vol. 11, no. 12, pp. 1466-14775, 2002.

- [24] J. H. Chang and J. M. M. Anderson, Regularized image reconstruction algorithms for positron emission tomography, IEEE Transactions on Medical Imaging, vol. 23, no. 9, pp. 1165-1175, 2004.
- [25] I. T. Hsiao, A. Rangarajan, and G. Gindi, A new convex edge-preserving median prior with applications to tomography, IEEE Transactions on Medical Imaging, vol. 22, no. 5, pp. 580-585, 2003.
- [26] Y. Chen, J. Ma, Q. Feng, L. Luo, P. Shi, and W. Chen, Nonlocal prior Bayesian tomographic reconstruction, Journal of Mathematical Imaging and Vision, vol. 30, pp. 133-146, 2008.
- [27] G. Sfikas, C. Nikou, N. Galatsanos, and C. Heinrich, Spatially varying mixtures incorporating line processes for image segmentation, Journal of Mathematical Imaging and Vision, vol. 36, no. 2, pp. 91-110, 2009.
- [28] S. Sanjay-Gopal and T. Hebert, Bayesian pixel classification using spatially variant finite mixtures and the generalized EM algorithm, IEEE Transactions on Image Processing, vol. 7, no. 7, pp. 1014-1028, 1998.
- [29] C. Nikou, N. Galatsanos, and A. Likas, A class-adaptive spatially variant mixture model for image segmentation, IEEE Transactions on Image Processing, vol. 16, no. 4, pp. 1121-1130, 2007.
- [30] C. Nikou, A. Likas, and N. Galatsanos, A Bayesian framework for image segmentation with spatially varying mixtures, IEEE Transactions on Image Processing, vol. 19, no. 9, pp. 2278-2289, 2010.
- [31] G. McLachlan, Finite Mixture Models, Wiley-Interscience, New York, 2000.
- [32] G. Sfikas and C. Nikou and N. Galatsanos, Edge preserving spatially varying mixtures for image segmentation, 2008.
- [33] S. Kotz, N. Balakrishnan, and N. L. Johnson, Continuous multivariate distributions, Wiley, New York, 2000.
- [34] H. R. Sheikh Z. Wang, A. C. Bovik and E. P. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Transactions on Image Processing, vol. 13, no. 4, pp. 600-612, 2004.
- [35] J. Nocedal and S. J. Wright, Numerical Optimization, Springer, Berlin, 1999.

Appendix

The experimental results of Chapter 4 are shown in more detail in this Appendix.

Algorithm	Execution time
MLEM	6 sec.
Gibbs	9 sec.
TV	33 sec.
GMM	54 sec.
GMM- CLP (EM)	267 sec.
GMM- $CLP (PCG)$	255 sec.
GMM-DLP (EM)	280 sec.
GMM-DLP (PCG)	268 sec.
GAMMA-MM	53 sec.
GAMMA-CLP (EM)	261 sec.
GAMMA-CLP (PCG)	241 sec.
GAMMA-DLP (EM)	277 sec.
GAMMA-DLP (PCG)	268 sec.

Table 5.1: Execution times for the compared algorithms (60 iterations, $\epsilon = 10^{-3}$).

ISNR				
Algorithm	75 photons	55 photons	$35 \mathrm{photons}$	$15 \ \mathrm{photons}$
MLEM	9.81	7.55	6.52	6.12
Gibbs	12.01	9.79	8.62	7.97
TV	15.08	11.09	9.42	8.58
GMM	15.10	11.65	10.07	9.30
GMM-CLP (EM)	17.40	12.62	10.63	9.59
GMM-CLP (PCG)	17.04	12.75	10.86	9.25
GMM-DLP (EM)	17.09	12.14	10.00	9.78
GMM-DLP (PCG)	17.18	12.73	10.19	9.99
GAMMA-MM	17.04	12.28	10.87	9.77
GAMMA-CLP (EM)	18.33	13.78	11.75	10.69
GAMMA-CLP (PCG)	18.57	13.80	11.75	10.74
GAMMA-DLP (EM)	18.20	13.89	11.70	10.69
GAMMA-DLP (PCG)	18.25	13.92	11.74	10.77

SSIM				
Algorithm	75 photons	$55 \ \mathrm{photons}$	$35 \mathrm{photons}$	$15 \ \mathrm{photons}$
MLEM	0.9973	0.9960	0.9953	0.9940
Gibbs	0.9999	0.9997	0.9974	0.9941
TV	1.000	0.9997	0.9975	0.9957
GMM	1.000	0.9998	0.9976	0.9961
GMM-CLP (EM)	1.000	0.9998	0.9976	0.9961
GMM-CLP (PCG)	1.000	0.9998	0.9977	0.9957
GMM-DLP (EM)	1.000	0.9997	0.9978	0.9962
GMM-DLP (PCG)	1.000	0.9997	0.9978	0.9960
GAMMA-MM	1.000	0.9998	0.9978	0.9964
GAMMA-CLP (EM)	1.000	0.9999	0.9979	0.9962
GAMMA-CLP (PCG)	1.000	0.9999	0.9978	0.9972
GAMMA-DLP (EM)	1.000	0.9999	0.9977	0.9962
GAMMA-DLP (PCG)	1.000	0.9999	0.9976	0.9962

		BIAS		
Algorithm	75 photons	55 photons	35 photons	15 photons
MLEM	$9 imes 10^{-5}$	$21 imes 10^{-5}$	$35 imes 10^{-5}$	44×10^{-5}
Gibbs	$8 imes 10^{-5}$	$21 imes 10^{-5}$	$34 imes 10^{-5}$	$43 imes 10^{-5}$
TV	$35 imes 10^{-6}$	$21 imes 10^{-5}$	$34 imes 10^{-5}$	$44 imes 10^{-5}$
GMM	$34 imes 10^{-6}$	$21 imes 10^{-5}$	$33 imes 10^{-5}$	$43 imes 10^{-5}$
GMM-CLP (EM)	$22 imes 10^{-6}$	$20 imes 10^{-5}$	$33 imes 10^{-5}$	$44 imes 10^{-5}$
GMM-CLP (PCG)	$18 imes 10^{-6}$	$21 imes 10^{-5}$	$32 imes 10^{-5}$	$46 imes 10^{-5}$
GMM-DLP (EM)	$18 imes 10^{-6}$	$20 imes 10^{-5}$	$30 imes 10^{-5}$	$43 imes 10^{-5}$
GMM-DLP (PCG)	$19 imes 10^{-6}$	$23 imes 10^{-5}$	$29 imes 10^{-5}$	$45 imes 10^{-5}$
GAMMA-MM	$20 imes 10^{-6}$	$19 imes 10^{-5}$	$29 imes 10^{-5}$	$43 imes 10^{-5}$
GAMMA-CLP (EM)	$20 imes 10^{-6}$	$20 imes 10^{-5}$	$27 imes10^{-5}$	$43 imes 10^{-5}$
GAMMA-CLP (PCG)	$17 imes 10^{-6}$	$20 imes 10^{-5}$	$27 imes 10^{-5}$	$44 imes 10^{-5}$
GAMMA-DLP (EM)	$18 imes 10^{-6}$	$20 imes 10^{-5}$	$26 imes 10^{-5}$	$43 imes 10^{-5}$
GAMMA-DLP (PCG)	$20 imes 10^{-6}$	$20 imes 10^{-5}$	$28 imes 10^{-5}$	$44 imes 10^{-5}$

VARIANCE					
Algorithm	$75 \ \mathrm{photons}$	55 photons	35 photons	15 photons	
MLEM	$18 imes 10^{-4}$	$25 imes 10^{-4}$	$59 imes 10^{-4}$	$96 imes 10^{-4}$	
Gibbs	$16 imes 10^{-4}$	$24 imes 10^{-4}$	$34 imes 10^{-4}$	$62 imes10^{-4}$	
TV	$9 imes 10^{-5}$	$13 imes 10^{-5}$	$19 imes 10^{-5}$	$29 imes 10^{-5}$	
GMM	$9 imes 10^{-5}$	$12 imes10^{-5}$	$17 imes 10^{-5}$	$28 imes 10^{-5}$	
GMM- CLP (EM)	$62 imes 10^{-6}$	$11 imes 10^{-5}$	$14 imes 10^{-5}$	$27 imes 10^{-5}$	
GMM-CLP (PCG)	$60 imes 10^{-6}$	$12 imes 10^{-5}$	$13 imes 10^{-5}$	$25 imes 10^{-5}$	
GMM-DLP (EM)	$64 imes 10^{-6}$	$12 imes 10^{-5}$	$16 imes 10^{-5}$	$27 imes 10^{-5}$	
GMM-DLP (PCG)	$63 imes 10^{-6}$	$11 imes 10^{-5}$	$17 imes10^{-5}$	$23 imes 10^{-5}$	
GAMMA-MM	$53 imes 10^{-6}$	$11 imes 10^{-5}$	$15 imes 10^{-5}$	$20 imes 10^{-5}$	
GAMMA-CLP (EM)	$52 imes 10^{-6}$	$11 imes 10^{-5}$	$16 imes 10^{-5}$	$25 imes 10^{-5}$	
GAMMA-CLP (PCG)	$51 imes 10^{-6}$	$12 imes 10^{-5}$	$14 imes 10^{-5}$	$22 imes 10^{-5}$	
GAMMA-DLP (EM)	$51 imes 10^{-6}$	$11 imes 10^{-5}$	$14 imes 10^{-5}$	$24 imes 10^{-5}$	
GAMMA-DLP (PCG)	$52 imes 10^{-6}$	$11 imes 10^{-5}$	$15 imes 10^{-5}$	$25 imes 10^{-5}$	

Table 5.2: Comparative statistics for various performance indices for the Shepp-Logan phantom. (a) ISNR (mean values of the 40 experiments), (b) structural similarity (mean value), (c) bias, (d) variance.

ISNR				
Algorithm	75 photons	55 photons	35 photons	$15 \ \mathrm{photons}$
MLEM	13.78	10.28	8.37	7.40
Gibbs	14.46	12.09	10.22	8.65
TV	18.44	12.41	10.35	9.34
GMM	18.45	12.51	10.39	9.50
GMM-CLP (EM)	18.99	12.66	10.39	9.53
GMM-CLP (PCG)	18.91	12.69	10.40	9.59
GMM-DLP (EM)	18.59	12.78	10.42	9.53
GMM-DLP (PCG)	18.79	12.65	10.46	9.51
GAMMA-MM	18.85	12.78	10.70	9.66
GAMMA-CLP (EM)	19.45	12.95	10.74	9.64
GAMMA-CLP (PCG)	19.75	12.92	10.79	9.62
GAMMA-DLP (EM)	19.61	12.96	10.74	9.62
GAMMA-DLP (PCG)	19.46	12.99	10.70	9.62

SSIM				
Algorithm	75 photons	$55 \ \mathrm{photons}$	$35 \mathrm{photons}$	$15 \ \mathrm{photons}$
MLEM	0.9998	0.9987	0.9968	0.9949
Gibbs	0.9999	0.9988	0.9970	0.9953
TV	1.000	0.9990	0.9971	0.9954
GMM	1.000	0.9991	0.9972	0.9955
GMM-CLP (EM)	1.000	0.9991	0.9972	0.9955
GMM-CLP (PCG)	1.000	0.9991	0.9972	0.9954
GMM-DLP (EM)	1.000	0.9991	0.9972	0.9955
GMM-DLP (PCG)	1.000	0.9991	0.9972	0.9954
GAMMA-MM	1.000	0.9991	0.9973	0.9956
GAMMA-CLP (EM)	1.000	0.9991	0.9973	0.9956
GAMMA-CLP (PCG)	1.000	0.9991	0.9973	0.9956
GAMMA-DLP (EM)	1.000	0.9991	0.9973	0.9957
GAMMA-DLP (PCG)	1.000	0.9991	0.9973	0.9956

		BIAS		
Algorithm	75 photons	55 photons	35 photons	15 photons
MLEM	$6 imes 10^{-5}$	$29 imes 10^{-5}$	$48 imes 10^{-5}$	61×10^{-5}
Gibbs	$3 imes 10^{-5}$	$29 imes 10^{-5}$	$48 imes 10^{-5}$	$60 imes 10^{-5}$
TV	$23 imes 10^{-6}$	$29 imes 10^{-5}$	$48 imes 10^{-5}$	$61 imes 10^{-5}$
GMM	$22 imes 10^{-6}$	$28 imes 10^{-5}$	$47 imes 10^{-5}$	$60 imes 10^{-5}$
GMM-CLP (EM)	$19 imes 10^{-6}$	$28 imes 10^{-5}$	$47 imes 10^{-5}$	$60 imes 10^{-5}$
GMM-CLP (PCG)	$12 imes 10^{-6}$	$28 imes 10^{-5}$	$47 imes 10^{-5}$	$60 imes 10^{-5}$
GMM-DLP (EM)	$12 imes 10^{-6}$	$28 imes 10^{-5}$	$47 imes 10^{-5}$	$60 imes 10^{-5}$
GMM-DLP (PCG)	$12 imes 10^{-6}$	$28 imes 10^{-5}$	$47 imes 10^{-5}$	$60 imes 10^{-5}$
GAMMA-MM	$17 imes10^{-6}$	$27 imes 10^{-5}$	$46 imes 10^{-5}$	$59 imes 10^{-5}$
GAMMA-CLP (EM)	$12 imes 10^{-6}$	$27 imes 10^{-5}$	$46 imes 10^{-5}$	$58 imes 10^{-5}$
GAMMA-CLP (PCG)	$11 imes 10^{-6}$	$27 imes 10^{-5}$	$46 imes 10^{-5}$	$59 imes 10^{-5}$
GAMMA-DLP (EM)	$11 imes 10^{-6}$	$28 imes 10^{-5}$	$46 imes 10^{-5}$	$59 imes 10^{-5}$
GAMMA-DLP (PCG)	$12 imes 10^{-6}$	$27 imes 10^{-5}$	$46 imes 10^{-5}$	$59 imes 10^{-5}$

VARIANCE					
Algorithm	$75 \ \mathrm{photons}$	55 photons	$35 \mathrm{photons}$	15 photons	
MLEM	$15 imes 10^{-4}$	$20 imes 10^{-4}$	$28 imes 10^{-4}$	56×10^{-4}	
Gibbs	$13 imes 10^{-4}$	$11 imes 10^{-4}$	$28 imes 10^{-4}$	$52 imes 10^{-4}$	
TV	$21 imes 10^{-6}$	$45 imes10^{-6}$	$26 imes 10^{-5}$	$27 imes10^{-5}$	
GMM	$20 imes 10^{-6}$	$44 imes 10^{-6}$	$25 imes10^{-5}$	$26 imes 10^{-5}$	
GMM- CLP (EM)	$17 imes 10^{-6}$	$36 imes 10^{-6}$	$24 imes10^{-5}$	$26 imes 10^{-5}$	
GMM-CLP (PCG)	$18 imes 10^{-6}$	$38 imes 10^{-6}$	$24 imes 10^{-5}$	$25 imes 10^{-5}$	
GMM- DLP (EM)	$21 imes 10^{-6}$	$32 imes 10^{-6}$	$25 imes 10^{-5}$	$26 imes 10^{-5}$	
GMM- $DLP (PCG)$	$20 imes 10^{-6}$	$32 imes 10^{-6}$	$24 imes 10^{-5}$	$25 imes 10^{-5}$	
GAMMA-MM	$16 imes 10^{-6}$	$35 imes10^{-6}$	$22 imes 10^{-5}$	$24 imes10^{-5}$	
GAMMA-CLP (EM)	$15 imes 10^{-6}$	$32 imes 10^{-6}$	$22 imes 10^{-5}$	$23 imes 10^{-5}$	
GAMMA-CLP (PCG)	$14 imes 10^{-6}$	$32 imes 10^{-6}$	$24 imes10^{-5}$	$22 imes 10^{-5}$	
GAMMA-DLP (EM)	$16 imes 10^{-6}$	$31 imes 10^{-6}$	$22 imes 10^{-5}$	$21 imes10^{-5}$	
GAMMA-DLP (PCG)	$17 imes10^{-6}$	$31 imes 10^{-6}$	$21 imes 10^{-5}$	$21 imes 10^{-5}$	

Table 5.3: Comparative statistics for various performance indices for the Elliptical phantom. (a) ISNR (mean values of the 40 experiments), (b) structural similarity (mean value), (c) bias, (d) variance.

Author's Publications

K. Papadimitriou and C. Nikou. Tomographic image reconstruction with a spatially varying Gaussian mixture prior. IEEE International Conference on Image Processing (ICIP'15), 27-30 September 2015, Quebec City, Canada.

Short Vita

Papadimitriou Katerina was born in Volos in 1987. She received her BSc degree (2011) in Computer Science from the Department of Computer Science and Engineering, University of Ioannina, Greece. In 2013, she became a postgraduate student at the same Institution. Her research interests lie in biomedical image processing, signal processing and statistical tomographic reconstruction methods.