Compressed Sensing for Few-View Multi-Pinhole Spect with Applications to Preclinical Imaging

Benjamin Michael Rizzo
Marquette University

COMPRESSED SENSING FOR FEW-VIEW MULTI-PINHOLE SPECT
WITH APPLICATIONS TO PRECLINICAL IMAGING

by

Benjamin M. Rizzo

A Dissertation submitted to the Faculty of the Graduate School,
Marquette University,
in Partial Fulfillment of the Requirements for
the Degree of Doctor of Philosophy

Milwaukee, Wisconsin

May 2018

ABSTRACT

COMPRESSED SENSING FOR FEW-VIEW MULTI-PINHOLE SPECT

WITH APPLICATIONS TO PRECLINICAL IMAGING

by

Benjamin Rizzo

Doctor of Philosophy in Computational Sciences

Marquette University, 2018

Single Photon Emission Computed Tomography (SPECT) can be used to identify and quantify changes in molecular and cellular targets involved in disease. A radiopharmaceutical that targets a specific metabolic function is administered to a subject and planar projections are formed by imaging emissions at different view angles around the subject. The reconstruction task is to determine the distribution of radioactivity within the subject from the projections. We present a reconstruction approach that utilizes only a few view angles, resulting in a highly underdetermined system, which could be used for dynamic imaging applications designed to quantify physiologic processes altered with disease.

We developed an approach to solving the underdetermined problem that incorporates a fast matrix-based multi-pinhole projection model into a primal-dual algorithm (Chambolle-Pock), tailored to perform penalized data fidelity minimization using the reconstruction's total variation as a sparse regularizer. The resulting algorithm was implemented on a Graphics Processing Unit (GPU), and validated by solving an idealized quadratic problem. Simulated noisy data from a digital rat thorax phantom was reconstructed using a range of regularizing parameters and primal-dual scale factors to control the smoothness of the reconstruction and the step-size in the iterative algorithm, respectively. The approach was characterized by evaluating data fidelity, convergence, and noise properties.

The proposed approach was then applied to few-view experimental data obtained in a preclinical SPECT study. ^{99m}Tc-labeled macroaggregated albumin (MAA) that accumulates in the lung was administered to a rat and multi-pinhole image data was acquired and reconstructed. The results demonstrate MAA uptake in the lungs is accurately quantified over a wide range of activity levels using as few as three view angles. In a pilot experiment, we also showed using 15 and 60 view angles that uptake of ^{99m}Tc-hexamethylpropyleneamineoxime in hyperoxia-exposed rats is higher than that in control rats, consistent with previous studies in our laboratory. Overall these experiments demonstrate the potential utility of the proposed method for few-view three-dimensional reconstruction of SPECT data for dynamic preclinical studies.

ACKNOWLEDGEMENTS

I would like to begin by expressing my gratitude and appreciation to the members of my committee for all the time, resources, and feedback that went into making this dissertation possible. In particular, I would like to thank my advisor, Dr. Anne Clough for investing a great deal her time (and patience) in me, and for her guidance throughout the course of this project. I would also like to thank Dr. Taly Gilat-Schmidt for providing access and support to the computing resources used in this dissertation, as all of the reconstruction and simulations presented in this dissertation were performed using her GPUs. I also owe a great deal of thanks to Dr. Said Audi. The protocols and the acquisition of the data presented in chapter 6 are in large part due to his assistance. I would also like to thank Dr. Emil Sidky for his advice and guidance with this project, as well as being quite generous with his time. Additionally, I would like to thank Dr. Sarah Hamilton and Dr. Daniel Rowe for their participation in the committee and preparing comments and suggestions.

I would like to also acknowledge the assistance from colleagues at Marquette, the Medical College of Wisconsin, and the Zablocki VA Medical Center. In particular I would like to thank Dr. George Corliss for his feedback on this dissertation, Steve Goodwin and Andy Williams for their technical support, Dana Scholler and Carlos Marquez for their assistance with experiments, and Dr. Elizabeth Jacobs for her feedback.

I would like to thank the department of Mathematics, Statistics and Computer Science at Marquette University for their support and the opportunities provided by the Richard W. Jobling fellowship, the Earl W. Swokowski fellowship, the Alvin and Marion

Finally, I would like to thank my family for their love and unwavering support. Thank you for always believing in me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ALGORITHMS

GLOSSARY

CC	Correlation coefficient
CP	Chambolle-Pock algorithm
D_{KL}	Kullback-Leibler divergence
$\Delta PPDG \times n$	Change in the partial primal dual gap multiplied by iteration
$E[x]$	Expectation of a distribution x
$f(x, y)$	Distribution of emissions
$\hat{f}(x, y)$	Estimate of the distribution of emissions
FOV	Field of view
FWHM	Full width at half maximum
g	Projection data
γ	Gamma ray photon
$\mathbf{H}$	Projection matrix
$\bar{\mathbf{H}}$	Single view projection matrix
h_{ij}	ij^{th} element of the projection matrix
$i2c$	Image to center-of-rotation distance
$i2p$	Image to pinhole distance
L_2	Euclidean distance (2-norm)
L	Largest singular value
λ	Smoothing parameter
M	Number of rows in the projection matrix (pixels)
MLEM	Maximum Likelihood Expectation Maximization

MPI	Message passing interface
N	Number of columns in the projection matrix (voxels)
NaI	Sodium iodide
OSEM	Ordered Subset Expectation Maximization
$p2c$	Pinhole to center-of-rotation distance
PMT	Photomultiplier tube
$PPDG$	Partial primal dual gap
$prox_\sigma[f](x)$	Proximal mapping of a function f
ϕ	View angle between -90º and 90º
RMSE	Root mean squared error
ROI	Region of interest
s	Primal-dual scaling parameter
SNR	Signal-to-noise ratio
σ	Dual step size
τ	Primal step size
θ	View angle
TcO_4^-	Pertechnetate
^{99m}Tc-MAA	Technetium-labeled macroaggregated albumin
^{99m}Tc-HMPAO	Technetium-labeled hexamethylpropyleneamineoxime
TV	Total variation
TV-KL	TV penalized Kullback-Leibler
TV-L$_2$	TV penalized L$_2$
μ	TV regularizer
ν	Scaling factor between the TV and data error metrics

Chapter 1

Introduction

Single photon emission computed tomography (SPECT) is a medical imaging

modality used to study the biological activity of tissues and organs noninvasively. A

pharmaceutical designed to target a specific physiologic function is labeled with a

radioactive isotope, administered to a patient, and allowed to distribute over time.

Photons emitted from the radiopharmaceutical are subsequently imaged using an image-

forming collimator coupled to a counting detector. Planar images taken from different

angles around the patient are acquired and used to reconstruct 3D digital volumes where

the intensity within each voxel is related to the amount of radioactivity emitted within the

imaged region of the subject. Thus, the distribution of the radiopharmaceutical can be

localized and quantified, and serves as an indicator of the activity of the targeted

metabolic process.

Clinical computed tomography, whether x-ray or emission, has traditionally relied

on analytic filtered back-projection methods to reconstruct digital volumes from a set of

2D planar images. Over the past twenty years, the Feldkamp method [32, 51] has formed

the foundation for most 3D reconstruction algorithms. These direct methods are fast and

mature, and when the projection image data is rich, i.e. a large number of views and

relatively high signal with low noise, they can produce exquisite reconstructions [13, 53].

However, due to practical patient limitations, SPECT imaging is often constrained to both

a suboptimal number of view angles to reduce the total imaging time, and to low activity

levels to minimize the radiation dose to the patient. Furthermore, the imaging hardware

required for SPECT has an inherently low detection efficiency, compared to x-ray CT, all contributing to a "photon-starved" imaging situation.

There has been much research within the past decade aimed at developing new approaches to SPECT imaging. These include i) development of new molecular agents that can be readily radiolabeled and have high specificity for the targeted cells, tissue, or organ of interest; ii) hardware developments such as modular cameras, faster and smaller electronics, pixelated detectors, optimal pinhole geometries (in the case of multi-pinhole SPECT); iii) novel iterative reconstruction methods designed to outperform traditional algorithms by incorporating the low count statistics property of the data and other physical properties of the SPECT system; and iv) the use of specialized computer hardware and software for computationally intensive tasks.

Statistical reconstruction methods, such as maximum likelihood expectation maximization (MLEM) [88] and ordered subset expectation maximization (OSEM)[47], have become standard reconstruction methods for SPECT [8, 68]. MLEM and OSEM algorithms account for the stochastic process of nuclear decay in the reconstruction model and hence provide better quality reconstructions from noisy data. Both MLEM and OSEM are iterative methods and involve multiple forward- and backward-projection operations until arriving at acceptable solutions. These methods can therefore be computationally intensive.

Within the past five years, new approaches to the reconstruction from projections problem have been developed to better reconstruct undersampled few-view CT data. These methods take advantage of two mathematical properties of the CT reconstruction problem; problems in computed tomography are typically convex, and algorithms can

exploit sparsity in some basis of the solution. Reconstruction methods can rely on existing convex optimization strategies, such as the Chambolle-Pock (CP) algorithm [17], while incorporating concepts of compressed sensing [15] as a prior.

Algorithms that utilize compressed sensing have been developed to solve underdetermined linear systems. Recently, the CP algorithm has been investigated in applications of few-view x-ray CT, where the goal is to reduce patient dose by reducing the number of view angles and the intensity of the x-ray source [91, 92, 93]. In this approach, the reconstruction methods strive to minimize the reconstructed volume's total variation (to yield "smooth" results) while insuring good agreement between the experimental data and the "estimated data" predicted from the reconstruction volume (data fidelity). Like the statistical methods, the CP algorithm is iterative. The reconstructed object is assumed to be approximately piecewise constant, and would therefore have a sparse gradient magnitude. However, unlike x-ray CT, in the SPECT problem the distribution of radiopharmaceuticals may be smoothly varying within the imaged volume thus requiring alternative approaches. Few-view SPECT using sparsity exploiting penalties has proven feasible in the case of 2D SPECT using the CP algorithm [104], and a sparsity exploiting approach to MLEM has been investigated in 3D SPECT [48].

SPECT systems for use with small animal studies have been developed for performing preclinical experiments in small laboratory animals aimed at identifying the mechanisms of disease, targets for treatments, and screening for development of drugs for such treatments [20, 80, 101]. SPECT is particularly attractive in the laboratory because of the relatively low price and complexity of a system, compared to other molecular

imaging modalities such as Magnetic Resonance Imaging (MRI) or Positron Emission Tomography (PET), and the range of molecular agents that can be labeled with Technetium, a commonly used gamma ray emitter. Our laboratory located at Zablocki VA Medical Center has designed and built a custom system optimized for studying lung physiology and pathology. In our system, like other animal systems, the image-formation collimator is either a single-pinhole collimator or, to acquire multiple views simultaneously, a multi-pinhole collimator. In our laboratory, the imaging task is to quantify the amount of activity within the lungs of a rat, and the focus is not on high resolution imaging.

^{99m}Tc labeled hexamethylpropyleneamine oxime (HMPAO), originally used to study brain perfusion [59, 60], is being used in our laboratory in preclinical studies as an early biomarker for pulmonary oxidative stress and sepsis in rats [6, 24, 28]. However accurate quantification of changes in the mechanisms associated with disease and reflected in HMPAO uptake, requires both a time-sequence of reconstructed image volumes and a pharmacokinetic model to interpret the resulting image data. The model accounts for the conversion of HMPAO between two different conformational forms in both the blood and lung tissue as well as other hemodynamic and physiologic factors. By fitting the model to the experimental SPECT data, the model could be used to estimate key parameter values that change with disease and hence provide targets for subsequent therapeutic interventions.

Techniques for performing few-view 3D SPECT reconstruction will provide the capability to perform dynamic pulmonary imaging studies in the future, and in combination with pharmacokinetic models provide a technology for noninvasive

measurements of physiological processes linked to lung injury. 3D dynamic SPECT studies would provide a means of quantifying the radiopharmaceutical uptake in the lung regions, and investigating the underlying mechanisms of lung disease/injury.

1.1 Objectives

Our laboratory at Zablocki VA Medical Center houses a unique micro-SPECT system designed for rat imaging. The overall objective of this research is to develop imaging strategies and reconstructions algorithms that will facilitate dynamic imaging of HMPAO, or other SPECT biomarkers, in lungs (or other organs) of small animals. SPECT imaging is our focus because imaging can be performed *in vivo* and is nondestructive, unlike other laboratory assays requiring lung excision, and produces volumetric maps of HMPAO activity, from which regional uptake of HMPAO can be quantified. Given this rat lung imaging problem, the algorithms and protocols we develop must address the following objectives:

- 3D volumetric imaging is required for accurate localization of the lung.

- The reconstruction algorithm must perform well with a relatively small number of view angles to reduce total acquisition time, since dynamic imaging will be used in the future.

- The reconstruction algorithm must produce images that are quantitatively accurate even at low activity levels, because uptake of HMPAO in the lungs is < 13% of injected dose.

- The reconstruction algorithm must be implemented so as to achieve reasonable computation times and storage requirements.

- The methods should be robust to experimental discrepancies or lack of precision in measurements and assumptions that occur in imaging of live laboratory animals.

The overall objective of this dissertation is to develop SPECT imaging algorithms, protocols, hardware, and software for rat lung imaging. To this end, we present the following specific aims.

1.1.1 Aim 1: Develop 3D SPECT reconstruction methods based on CP.

Development of 3D imaging methods are required to accurately identify the lung regions in the reconstructed image volume. If the methods are to be used to reconstruct dynamic data, the methods must be fast, and an algorithm capable of reconstructing limited view angle data is required. Thus, Aim 1 consists of two tasks.

i) The first task will be to develop a fast rotation-based matrix-projection model for 3D SPECT reconstruction, implemented on a GPU.

ii) The second task is to implement a Chambolle-Pock (CP) -type reconstruction method that utilizes the rotation-based matrix-projection model.

1.1.2 Aim 2: Validate and characterize the reconstruction methods.

Methods developed in Aim 1 will be tested and evaluated on phantom data to verify the forward- and back-projectors are well matched and to characterize the convergence behavior of the few-view 3D SPECT reconstruction algorithm. Furthermore, the phantom reconstructions will be used to evaluate the quantitative accuracy of the

methods over a range in simulated activity levels, as would be required for a dynamic study.

i) The first task of this objective will be to validate the CP method using our GPU implementation by reconstructing a simple idealized phantom. Analysis of the CP convergence metrics applied to the "quadratic-roughness" problem and the penalized Kullback-Leibler problem will be used to confirm the suitability of our rotation-based approach.

ii) Secondly, optimal design parameters for the micro-SPECT system will be identified. Analysis and simulations will be used to optimize the design parameters of the algorithm, with a focus on its use with data from our micro-SPECT system. The methods will be developed and tested using digital rat thorax phantoms and experimental rat data. The MLEM reconstruction algorithm will be implemented on the GPU to serve as a comparison and validation method.

1.1.3 Aim 3: Demonstrate the feasibility on preclinical SPECT data.

The final objective is to extend the methods to experimental rat data obtained in our laboratory. Two pulmonary image experiments will be performed *in vivo* to test how well the methods can localize the lungs, and to test if the methods can be used to detect a difference in the uptake of ^{99m}Tc-HMPAO between control and hyperoxic exposed rats. Reconstructions of the experimental data will be used to further refine the 3D projection model and evaluate the CP method under few-view and low-dose conditions, as would be expected in a dynamic study.

1.2 Chapter summaries

     Chapter 2 provides background for SPECT imaging, including an introduction to the forward- and back-projection operations, and SPECT acquisition.

     Chapter 3 describes a rotation-based approach to iterative reconstruction using fast matrix multiplication methods and GPUs.

     Chapter 4 provides an overview of direct and iterative methods for SPECT. The theory and the development of MLEM and the CP reconstruction algorithms is presented.

     Chapter 5 validates the proposed rotation-based system matrix approach to the projection model applied to CP reconstructions using a simple idealized phantom. This chapter also characterizes and evaluates the convergence behavior of the CP method applied to gradient magnitude penalized Kullback-Leibler minimization for reconstructing a digital phantom of a rat thorax.

     Chapter 6 presents reconstructions of experimental SPECT data used to evaluate the suitability of the CP algorithm *in vivo* pre-clinical studies.

     Chapter 7 summarizes the results of this dissertation, provides suggestions for future work, and concludes the dissertation.

Chapter 2

Background

2.1 SPECT imaging

The objective of SPECT imaging is to determine the distribution of γ emission activity within a 3D volume. The process begins with the administration of a gamma-emitting radiopharmaceutical that localizes within particular tissues, depending on its molecular composition. The challenge is then to adequately detect the emissions and determine the position of their origination within the object.

The underlying principles of SPECT are based on principles of x-ray computed tomography, ideas first proposed by Radon [81] and developed by Cormack and Hounsfeld [27, 46]. In the case of SPECT, photons originating from within the subject are imaged using some type of a gamma camera, for example, a position-sensitive gamma detector coupled with an imaging collimator. An image acquired by the gamma camera is known as a *projection* of the emission distribution, where the image is formed by counting the number of γ photons incident on each detector element. Data obtained from a single projection of the object contains no information about the depth within the volume at which a given photon originated. Projections from multiple view angles are therefore acquired in order to localize in 3D the position of each recorded gamma emission. Generally, more projections provide better information about the relative positions of the emissions within the object volume.

2.1.1 The forward-projection and back-projection operation

We begin by ignoring any interactions of an emitted photon within the object during its path to the detector. Then projections may be modeled mathematically by the Radon transform [53], as illustrated in 2D in Figure 2.1. The Radon transform, $g(t, \theta)$ (projection data) of an emission distribution, $f(x, y)$ (unknown object), is the sum of all values of $f(x, y)$ along the rays defined by the Dirac delta function, $\delta(x\cos\theta + y\sin\theta - t)$. Figure 2.1 depicts a "parallel hole" collimator, where photons emitted only along the lines $t = x\cos\theta + y\sin\theta$ are counted. In general, the kernel given here by the delta function is specific to the particular imaging collimator used.

$$g(t, \theta) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y)\, \delta(x\cos\theta + y\sin\theta - t)\, dx\, dy$$
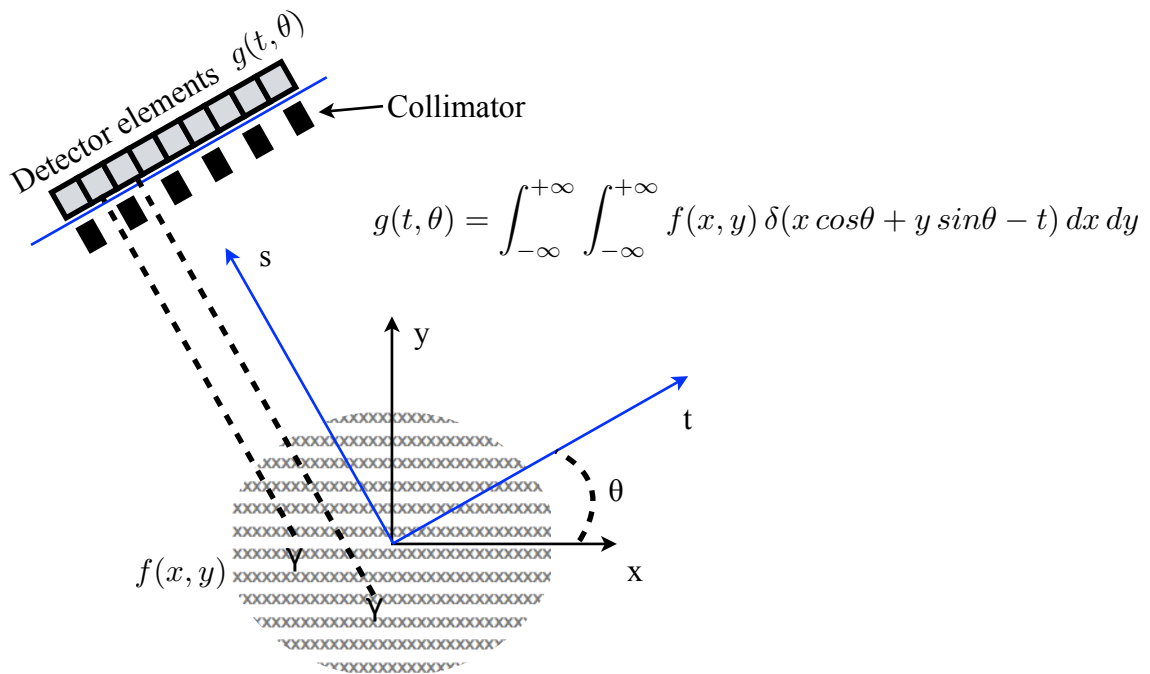
Figure 2.1: The Radon transform (forward-projection) for a parallel imaging collimator.

In practice, the data collected $g(t, \theta)$ is a discrete set of pixel intensities corresponding to the counts acquired by the detector elements at grid positions (t, θ), and the distribution f is a discrete set of voxel intensities at grid positions (x, y) in the imaged

plane representing the distributions of emissions. In its discrete form, the value of g at an individual detector element, g_i, is the weighted sum of all grid elements f_j,

$$g_i = \sum_{j=0}^{N-1} h_{ij} f_j \tag{2.1}$$

where N is the number of grid elements and the weight, h_{ij}, represents the probability that a photon originating from the jth grid element is detected by the ith detector element. Like the delta function, the values of the weights depend on the image-formation collimator used in the imaging system and the view angle. The weights may also be used to incorporate physical models such as photon attenuation and scatter along the emission-to-detector path [25] and/or non-ideal collimator properties [102].

Equation 2.1 can be extended easily to 3D SPECT systems, where the set of all g_i for a single view angle represents the pixel intensities of a planar image, and f_j is the set of all voxel intensities at grid locations (x, y, z) in the reconstruction volume. More compactly, the entire forward projection may be written in matrix form as,

$$g = \mathbf{H} f \tag{2.2}$$

where $\mathbf{H}$ is the forward projection matrix composed of elements h_{ij}. The elements of vector g consist of all pixel intensities from all view-angle projections, and the elements of the vector f represent the voxel intensities of the imaged volume. Thus the reconstruction problem in SPECT may be thought of as determining the distribution of emissions f from the counts acquired over all projections g.

Likewise, a mapping from a discrete set of pixel coordinates $g(t, \theta)$ to voxels positions (x, y) can be defined using the conjugate of the forward-projection, know as the

back-projection. As in the example of Figure 2.1, Figure 2.2 illustrates the back-projection operation for a single view angle. Counts from each detector element are reassigned to points (x, y) that fall along the ray defined by $t = x\, cos\theta + y\, sin\theta$, for a parallel collimator model. The superposition of all back-projections from every view

$$\hat{f} = \int_0^\pi g(t, \theta)|_{t=x\cos\theta+y\sin\theta}\, d\theta$$
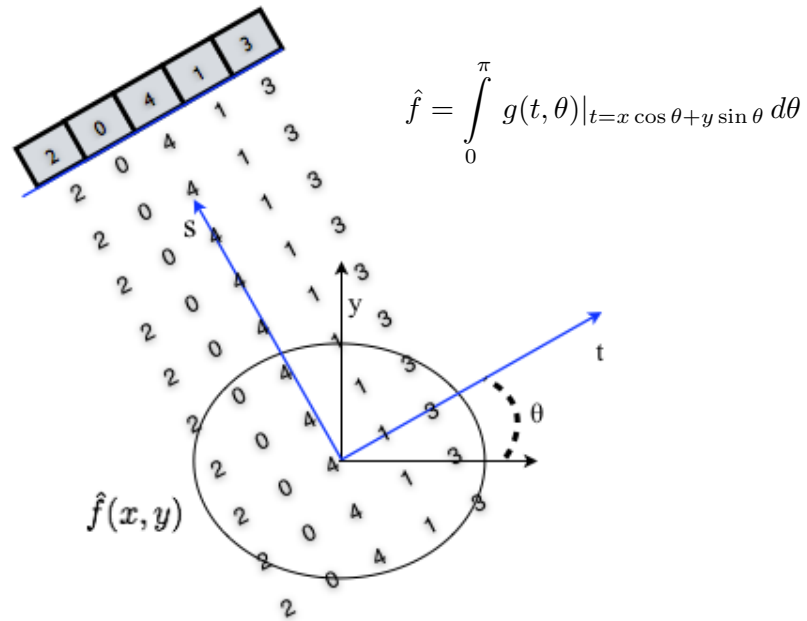
Figure 2.2: The back-projection operation for a single view angle with a parallel hole collimator model.

angle adds up to an intensity map $\hat{f}$ onto the reconstruction space. In its discrete form, the back-projection is defined as

$$\hat{f}_j = \sum_{i=0}^{M-1} h_{ji}\, g_i \, .$$

(2.3)

This implies that the back-projection operation may be carried out using the transpose of **H** from Equation 2.2,

$$\hat{f} = \mathbf{H}^T g \, .$$

(2.4)

In general $\mathbf{H}$ is not a square matrix, the data vector g contains noise, and the dimensions of all components can be quite large, so that the solution of Equation 2.2 is not straightforward. Furthermore, Equation 2.2 represents a highly undetermined system, in that the number of rows in $\mathbf{H}$ is generally substantially less than the number of columns, for which a unique solution does not exist. The forward- and back-projectors are essential operations in many approaches to the reconstruction problem, i.e. determining f from g. Methods for generating good approximations to the true emission distribution based on forward- and back-projection operations for different imaging systems, each corresponding to a different $\mathbf{H}$ matrix, are discussed in Chapter 4.

2.2 SPECT image formation

A SPECT imaging study begins with administration of a radiopharmaceutical to a subject. Commonly this involves an intravenous injection of the radiopharmaceutical which then distributes throughout the body via the bloodstream. After allowing time for the radiopharmaceutical to be taken up within the targeted tissue, data is acquired over a fixed time period that is long enough to accumulate a sufficient number of photon counts to obtain a statistically reliable projection image at each view-angle. The exact clinical protocol depends on the activity of the radioisotope as well as the clinical aims of the imaging study.

SPECT projections are formed using image collimators. The image is formed much like a camera lens system, where light photons originating from the object are focused into an image according to the geometry of the lens or lens system. However, radioisotopes used in SPECT imaging, such as ^{99m}Tc, emit gamma photons which cannot

be redirected using refraction, as in an optical system, and cannot be detected directly

using conventional photosensitive electronics. Instead, SPECT images are generally

formed by using collimators containing small aperture(s) bored into a piece of lead or

tungsten. Each aperture can only accept photons traveling along rays defined by the

collimator geometry. Some common ideal image forming collimators are shown in figure

2.3, with pinhole geometries that are widely employed for tomographic studies.

     As shown in figure 2.3, pinhole collimators only accept gamma rays which pass

through a single point to form a projection of the object, while all other gamma rays are

blocked by the walls of the collimator. Like an optical system, how well focused the

image is depends on the size of the focal point, i.e. the diameter of the pinhole in the case

of pinhole SPECT. Also like optical systems, the magnification of the image can be

changed by adjusting the distance between the image and focal point, i.e. the distance

between the object and the focal point.
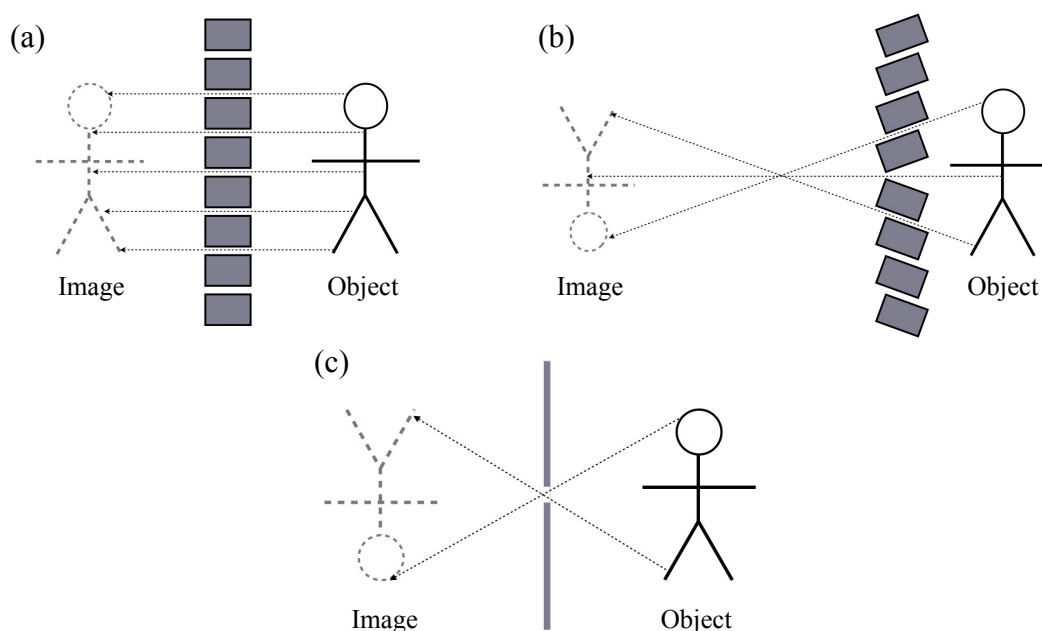
Figure 2.3: Ideal collimators: (a) parallel beam, (b) fan beam, and (c) pinhole collimator.

Pinhole collimators are not typically employed in traditional optical systems due to their small solid angle of acceptance, which only lets in a limited number of photons originating from the subject. Unless there is a high rate of photon detections, pinholes require long exposure times to form adequate images using optical camera systems, and this is also the case for pinhole SPECT. In a pinhole SPECT imaging system trade-offs between acquisition time, patient dose, and image focus must be considered. Acquisition times may be reduced by increasing the amount of radioactivity used in the imaging protocol, but this comes at the cost of a higher dose to the subject. Alternatively, acquisition times may be reduced by increasing the pinhole diameter, which provides a larger angle of acceptance, and allows more photons to pass through the collimator. However, this comes at the cost of increased blur in the projections as illustrated in figure 2.4(a), and subsequent blurring in the reconstructions, reducing the system's ability to resolve small features in the projections and reconstructions. This blurring can be incorporated into the projection model by measuring the photon intensity profile for a point source of activity. Functionally, this is known as the point-spread function of the pinhole, illustrated in figure 2.4(b). Methods for modeling the pinhole blur are discussed in section 2.3, while the photon counting statistics of the detection process are discussed in Chapter 4.

(a)   (b)
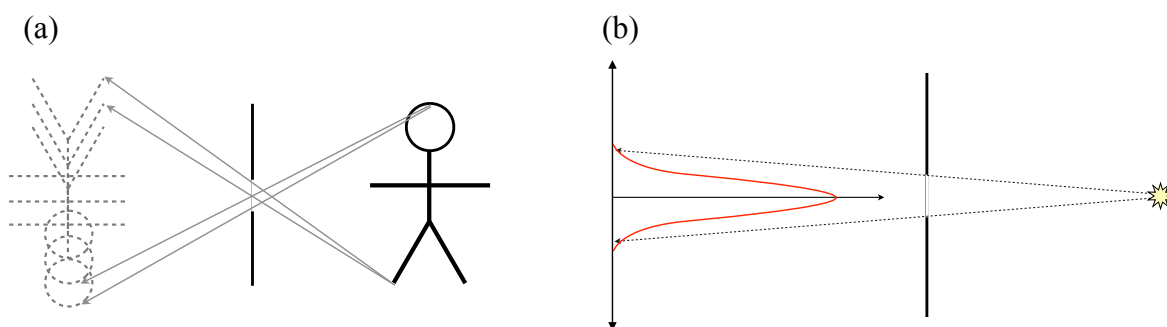
Figure 2.4: Non-ideal pinhole collimator: (a) blur induced by the finite diameter of the pinhole, and (b) the pinhole's point-response function.

In order to process the image formed by the collimator, the gamma ray photons must be converted into an electrical signal. Using a scintillation camera [3, 9, 34], a signal can be acquired using a sodium iodide crystal (NaI) coupled to an array of photo multiplier tubes (PMT). The gamma cameras used to collect data for this dissertation each consist of a 3 x 3 PMT array designed to the specifications of FastSPECT II [34], and provided to us by the Center for Gamma Ray Imaging at the University of Arizona under the direction of Harrison Barrett, Ph.D.

The detection and conversion process is depicted in figure 2.5. A gamma photon that is incident on the face of a detector interacts with the lattice of a thin layer of NaI crystal doped with thallium. Visible photons are emitted from the point of interaction between the NaI crystal, and travel down a quartz light guide. The visible photons are converted to an electronic signal using PMTs. Visible photons that are incident on the face of the PMT collide with free electrons on a photocathode. The electrons are freed from the surface of the photocathode and accelerated through a series of plates, or dynodes, set at progressively higher voltages. After colliding with a dynode plate, secondary electrons are released and accelerated towards the next dynode. In this way, a single electron from the photocathode is amplified into a measurable output current. Fast electronics are then used to analyze the outputs from each PMT and the outputs are saved in list-mode format [9, 78]. The projection image is then calculated from the list of 9 PMT outputs as discussed in Appendix A.

17
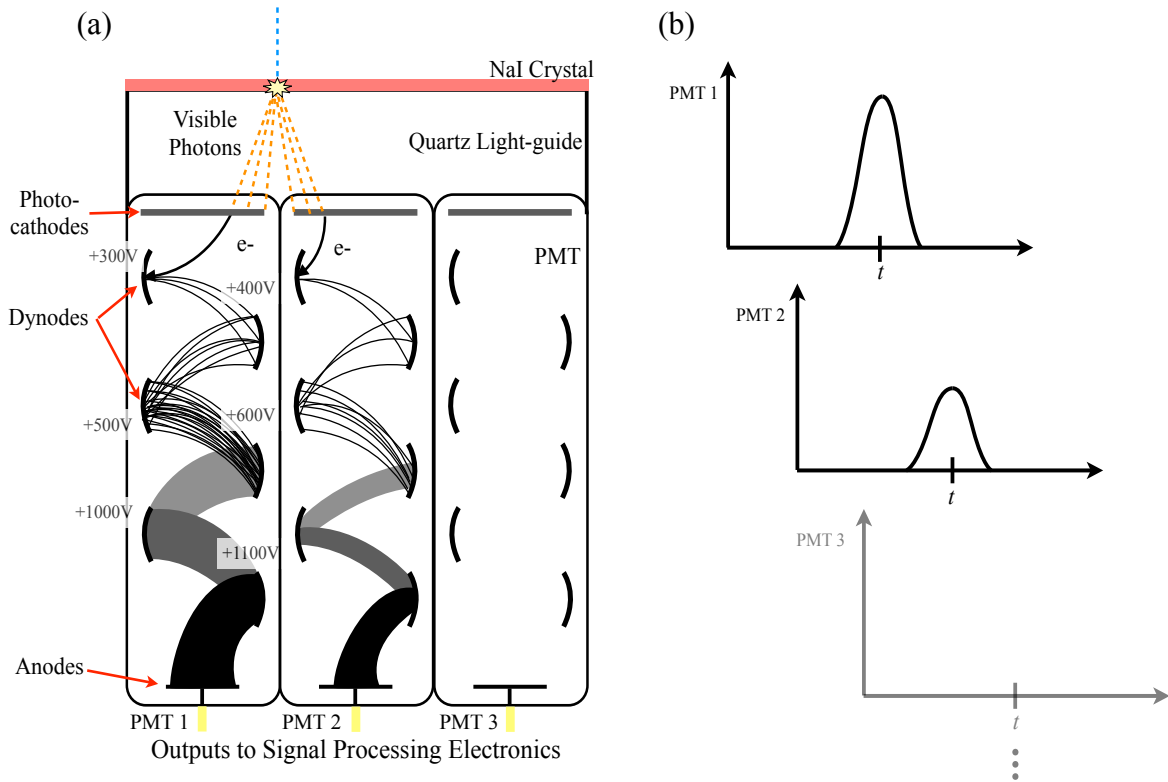
Figure 2.5: (a) Cross-section of a NaI gamma camera depicting a detection event (modified from [14]), and (b) each PMT's response to the event, where only 3 of the 9 outputs are depicted.

2.3 The SPECT system matrix

The forward- and back-projection operations in Equations 2.2 and 2.3 both require the system matrix, **H**, that models the geometry of the collimator. In order to calculate **H**, a 3D grid is imposed on the reconstruction volume, consisting of equally spaced and orthogonal parallel planes. The voxels are then defined to be the cube elements inside the intersecting planes. Likewise, a 2D grid of pixels makes up the detector plane of the gamma camera. The matrix elements of **H** represent the probability that an emission in the jth voxel element is detected in the ith detector element, and are calculated geometrically by tracing a ray between detector pixels and the reconstruction voxels, as shown in figure 2.6.
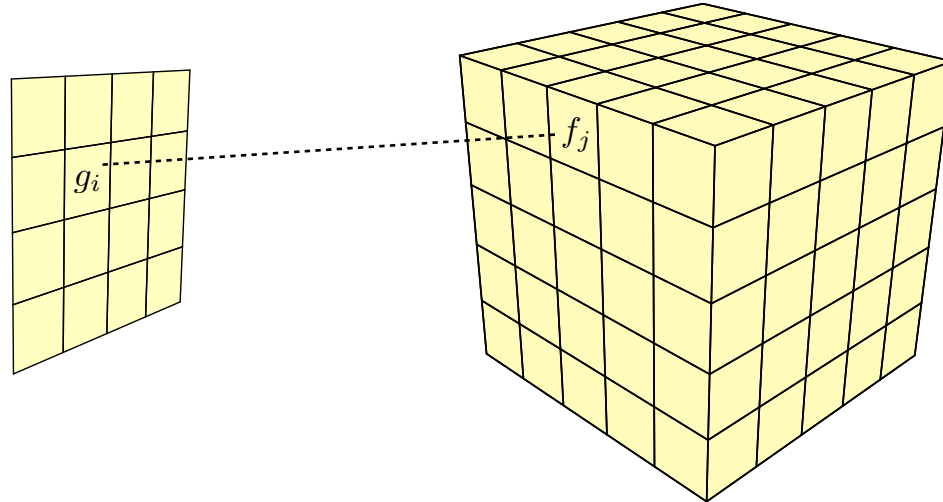
Figure 2.6: A discretized reconstruction volume and detector plane.

A fast method for ray tracing, proposed by Siddon [90] and improved by Han [40], calculates the elements of **H** by modeling the ray traced between detector and reconstruction coordinates as a parametric line. Determining the location of the intersections between the line and the parallel planes of the reconstruction volume can be easily calculated. The matrix weights may then be calculated using the length of the intersections compared to the total length of the ray, as shown in figure 2.7, where the image-to-center of rotation distance is given by *i2c* and the image-to-pinhole distance is given by *i2p*. Because of its simplicity, Siddon's method may be easily adapted to any imaging geometry, though a pinhole geometry is depicted in the figure and is assumed in the proceeding discussion.

There are two standard methods for calculating **H**, a "voxel-driven" and a "ray-driven" approach. In the ray-driven approach, illustrated in figure 2.7, rays are traced from individual detector elements through the center of the pinhole opening. Voxels along the path of the ray are used to calculate the intersection weights for the corresponding
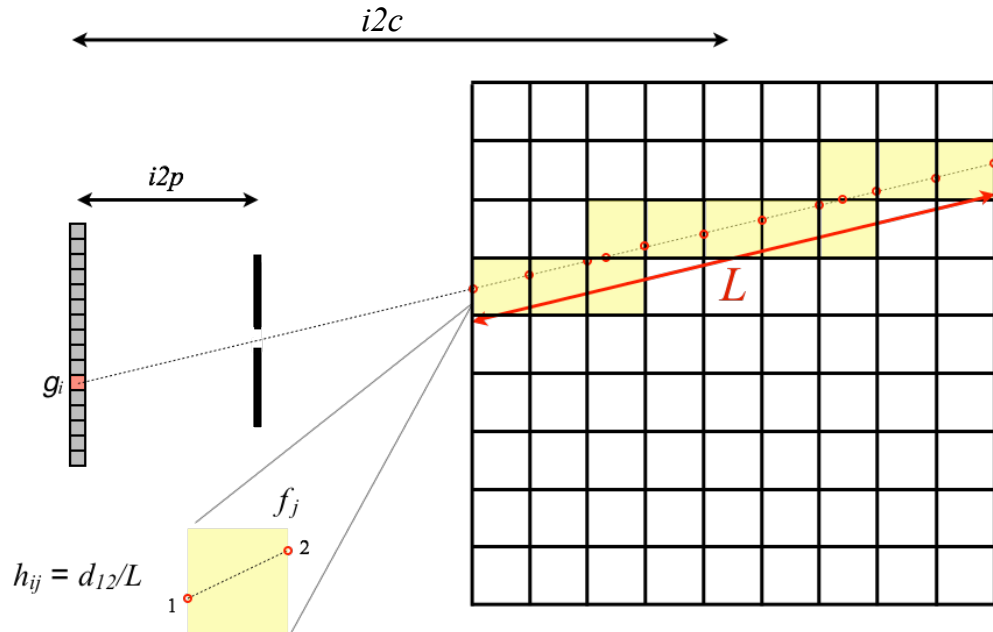
Figure 2.7: Illustration of Siddon's ray tracing method for calculating matrix weights, using a ray-driven approach.

pixel. In the voxel-driven approach, illustrated in figure 2.8, rays are traced from

individual voxels through the center of the pinhole. Matrix weights are then determined
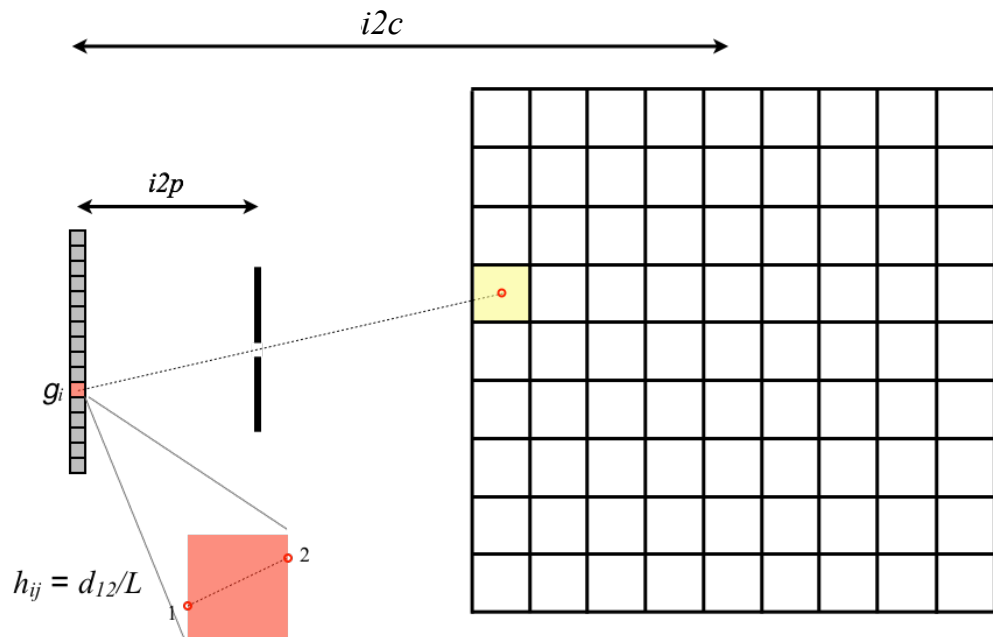
Figure 2.8: Voxel-driven ray tracing approach.

using each intersecting pixel.

There are advantages and disadvantages to each approach. The ray-driven approach involves iterating over all M detector elements, whereas the voxel-driven approach involves iterating over all N voxel elements. Since M is generally less than N, the ray-driven method is less computationally expensive than the voxel-driven method. Additionally, each method models the forward and back-projection operation differently. Figure 2.9 shows an illustration of the two methods using a single point to model the
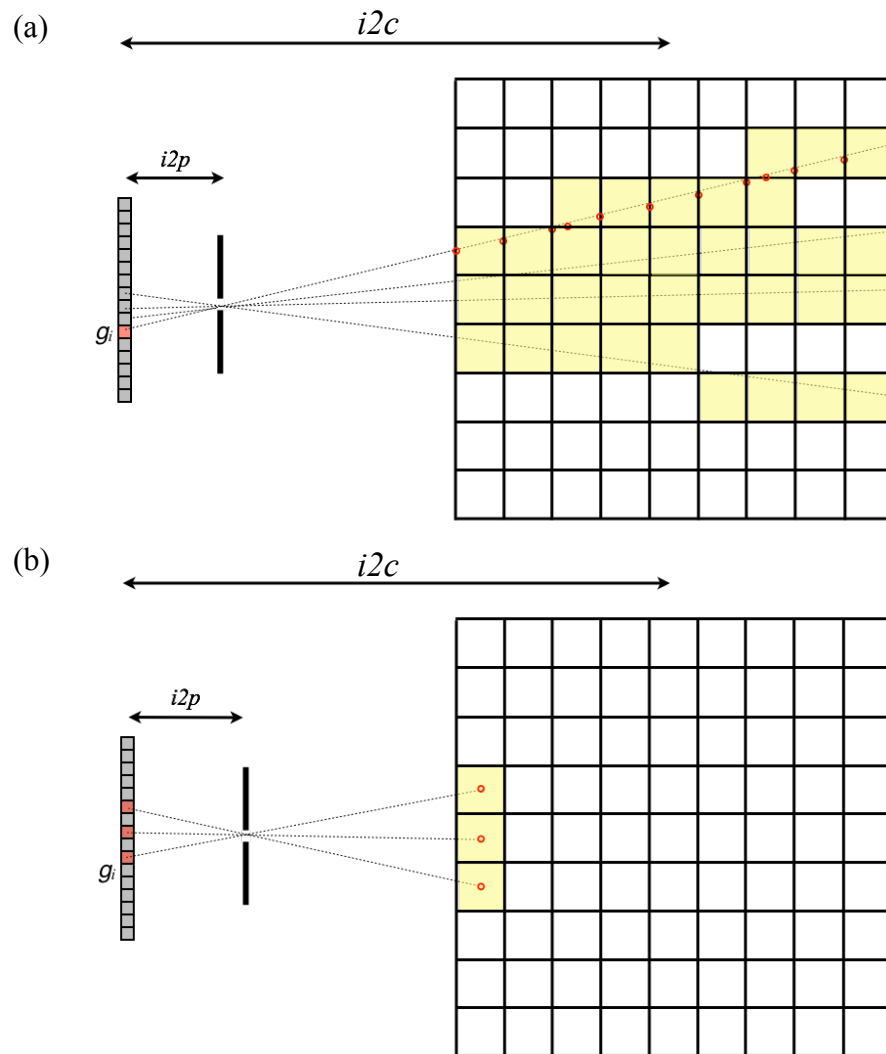
Figure 2.9: (a) Ray-driven approach, where some voxels do not fall within the path of rays traced from the detector. (b) Voxel-driven approach, where some pixels do not fall within the path of rays traced from the reconstruction space.

pinhole. In the ray tracing method, each pixel is mapped to voxels in the reconstruction space. Only voxels that fall along the line that intersects the detector element and the pinhole center are included in h_{ij}. This can lead to gaps in coverage in the reconstruction space, where the contribution of select voxels are missing in the system matrix. Likewise, in the voxel method, each voxel is mapped to the detector space. Only detector elements that fall within the path of intersection are included in h_{ij}. Hence, the contribution of select pixels are missing in the system matrix.

In x-ray CT, where the focal point is generally much smaller than the voxel (μ m vs mm), one approach to account for the gaps in coverage in the ray tracing model is to use the ray-driven model for the forward-projection, and the voxel-driven model for the back-projection. Reconstructions using this approach are said to have mismatched forward and back-projector pairs [106]. The unmatched pair approach has been used to better model photon attenuation in SPECT [62, 106], and can be used as a strategy to accelerate the rate of convergence of iterative reconstruction methods [57].

In SPECT, where the dimension of the pinhole diameter is on the same order of magnitude as the pixel spacing, a common approach is to use a voxel-driven method in conjunction with a blurring operation, typically Gaussian, that models the point-response of the pinhole. Blurring may be thought of as a mixing of weights from neighboring pixels, reducing the gaps in coverage inherent in the voxel-driven method. The full forward-projection in the voxel-driven approach is then the product of the geometric weights, **H**, with a blurring matrix **B**,

$$\mathbf{H_B} f = \mathbf{BH} f = g \, .$$

(2.4)

A more recent approach in SPECT incorporates the size of the pinhole into a ray-driven model [102]. This is accomplished by modeling the pinhole as a collection of points that fall within the diameter of the pinhole, as in figure 2.10(b), as opposed to using a single point, as in figure 2.10(a). Multiple rays are traced from an individual detector element, one for each point within the pinhole diameter, which reduces the number of voxels missing in the system matrix. This approach was demonstrated using single pinhole SPECT [102], which allows for additional physical aspects to be incorporated into the projection model, such as photon attenuation and pinhole edge penetration [95]. For example, the opening and the walls of the pinhole may be modeled as a mask, as shown below in figure 2.10(b), where the probability of passing through the opening is 1 and the probability of passing through the wall is zero. The probability of a photon passing through the walls of the pinhole may then be modeled by re-weighting the probabilities around the edges of the pinhole opening [102], as in figure 2.10(c).

For this dissertation the system matrix, $\mathbf{H}$, is modeled using the ray-driven method proposed by Wietholt [102], and the pinhole mask in figure 2.10(b). This method was chosen because it is faster than a voxel-driven approach, and automatically incorporates pinhole blurring. Additionally, we are constrained to use well matched forward- and back-projector pairs in the reconstruction method developed subsequently in chapter 3. i.e, we are required to use $\mathbf{H}$ for the forward-projection and $\mathbf{H}^{T'}$ for the back-projection operators. The ray-driven pinhole mask approach will be extended to a multi-pinhole projection model, and will be demonstrated on fully 3D SPECT data using a compressed sensing algorithm.
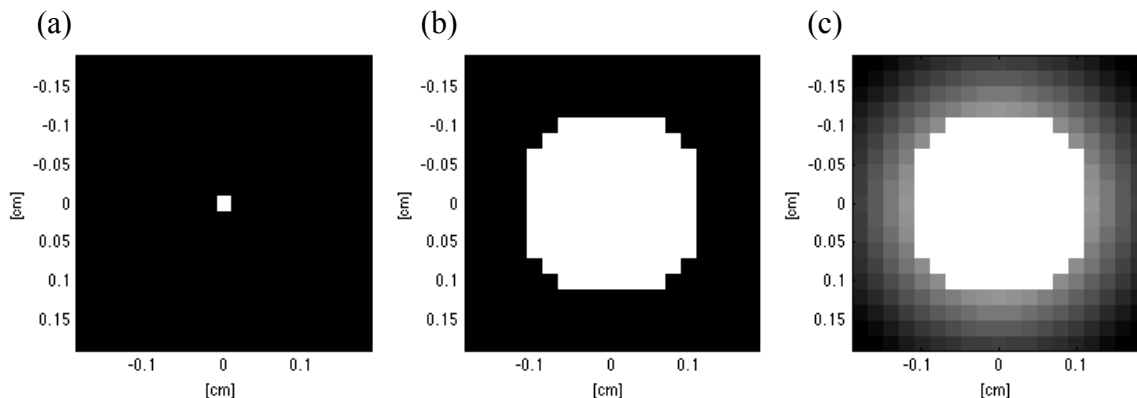
Figure 2.10: Pinhole models; (a) Ideal pinhole, (b) 2 mm pinhole, and (c) 2 mm pinhole with edge penetration.

2.4 SPECT imaging studies and protocols

The protocols used in SPECT imaging studies vary depending on the targeted organ/tissue and the type of information sought in the study. In all SPECT studies, a radiopharmaceutical that targets a specific tissue/organ of interest is administered to the patient. In a static study, a period of time elapses prior to imaging which permits a steady state distribution of emissions to be reached so that there is no appreciable change during the acquisition process. Many individual projections are then taken either using a single gamma camera by rotating the camera around the patient, or state-of-the-art clinical scanners typically use multiple gamma cameras rotated around the patient to reduce the overall scan time.

For studies aimed at determining metabolic function or dysfunction, it is often necessary to acquire dynamic images to determine how the distribution of the radiopharmaceutical changes with time. Thus, a time sequence of projections is acquired simultaneously at multiple view angles around the subject and an image volume is reconstructed at each time point. This requires the use of stationary SPECT systems,

where multiple gamma cameras are placed at fixed view angles around the patient/subject to simultaneously acquire projection data with each camera. If the rate of the radiopharmaceutical kinetics is not too big, it may also be feasible to use a multiple gamma camera system and allow the cameras to rotate over only a few view angles to obtain more total views. Experimental time-activity-curves (TAC) are obtained by determining the mean activity within a volume/tissue of interest in each reconstructed volume. Pharmacokinetic analysis, which models how the pharmaceutical interacts with tissue(s), can then be used to identify targeted molecular and cellular changes in the organ or tissue of interest, and can be a tool in identifying disease/injury.

In the case of small animal preclinical SPECT (or micro-SPECT), similar static and dynamic protocols may be used to study physiology and disease pathology, where the system's geometry, design, and reconstruction methods are generally chosen depending on the types of studies being performed. Static studies may be carried out by rotating the gamma camera(s) around the specimen over many view angles, resulting in a data rich acquisition. These systems can better resolve the distribution of emissions and can employ standard CT reconstruction methods, such as filtered back-projection (FBP) (see chapter 4.1.1). Since the features being imaged in a micro-SPECT system are small, pinhole collimators, as in figure 2.11, can be used as a means of controlling the magnification of the projections. Smaller features in the emission distributions may be resolved using higher magnification and smaller pinholes, and can be used to compliment data rich acquisitions.

Stationary preclinical SPECT systems are typically constrained to fewer than 10 angles due to the size of the detectors and the cost and complexity of multi-camera
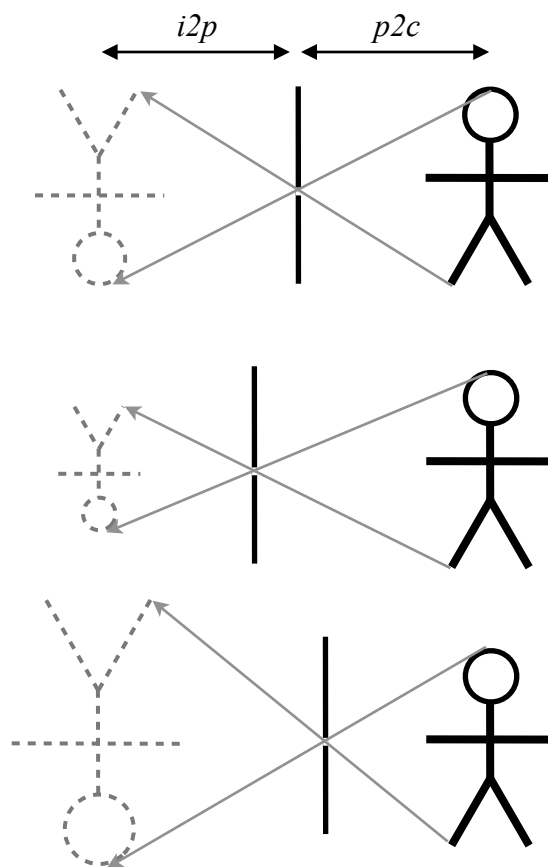
Figure 2.11: Adjusting the magnification of pinhole collimators by changing the image-to-pinhole (*i2p*)/ pinhole-to-center of rotation (*p2c*).

systems, resulting in *limited angle* data. For example, the FastSPECT II [8, 34] system, developed by the Center for Gamma Ray Imaging at the University of Arizona, employs two rings of 8 (5" x 5") stationary NaI gamma cameras [54], pictured in figure 2.12, to cover the entire animal.

The U-SPECT II system, developed by MILabs, employs 3 large NaI detectors (20" x 15")[10, 100] to obtain large magnifications as illustrated in figure 2.13. Both systems acquire far fewer projections than the several hundred required by FBP methods. This limited number of views negatively impacts the quality of the reconstructed images due to lack of sufficient information within the data to adequately determine the location

Figure 2.12: Picture of the FastSPECT II system with the outer shielding removed, revealing the gamma camera arrangement [54].
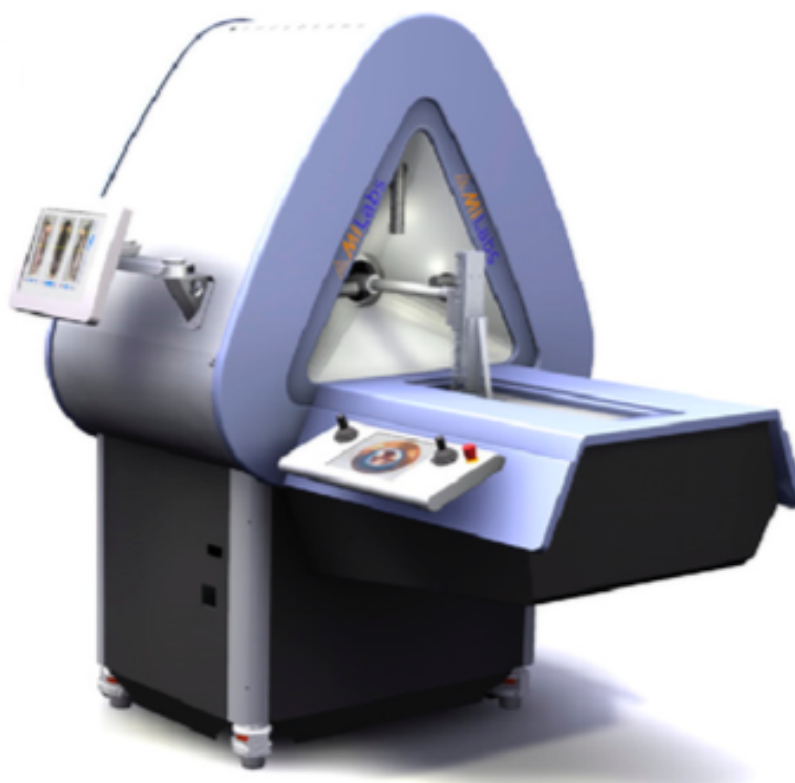
Figure 2.13: U-SPECT II imaging system (figure 1(A) in [100]).

of a given photon emission. For systems with limited view angles, multiple pinhole collimators are often used to provide more information per view, provide a larger field-of-view (FOV), increase the signal-to-noise ratio (SNR) [65], and provide better spatial resolution [10]. FastSPECT II uses a system of adaptive multi-pinhole collimators [7, 34], while U-SPECT II uses 25 pinhole projections per camera [10, 100]. With the combination of large magnification and multi-pinhole collimators U-SPECT II is capable of mouse brain studies, with resolution nearing 0.35 mm. Both systems are designed to obtain data with geometries that maximize the amount of angular coverage.

Our system is a custom designed and built micro-SPECT system at the Zablocki VA Medical Center, Milwaukee WI. A schematic of the overall experimental setup is shown in figure 2.14 and is designed for studying lung injury models in rats, with the long term goal of dynamic imaging studies. The aims of a dynamic study are on activity quantification and since the lungs are a relatively large organ, high resolution images are
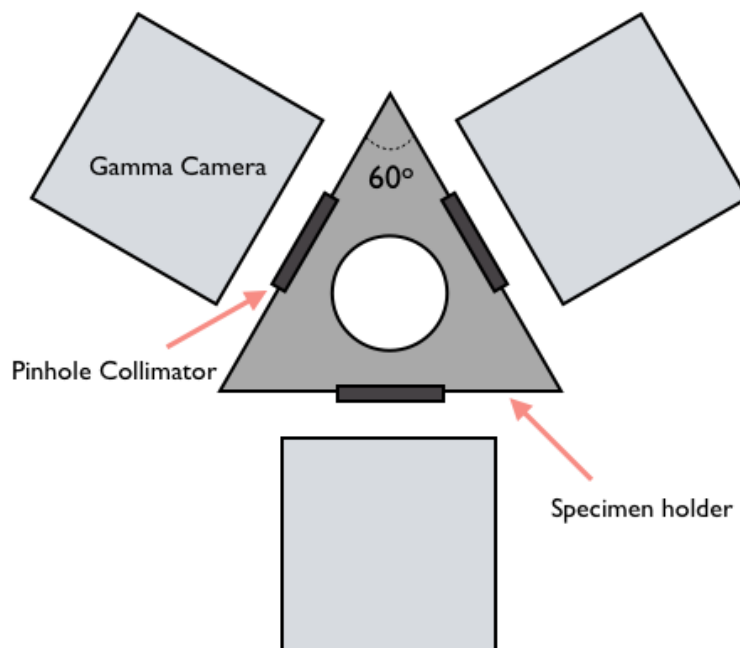
Figure 2.14: Schematic of the micro-SPECT system.

highly desirable but are not the primary objective. The system is designed to use 3 small stationary modular gamma cameras spaced 120º apart and placed around a small animal specimen holder, where the cameras and data acquisition electronics are designed to the FastSPECT II specifications [34].

Our SPECT system uses a five pinhole lead collimator pictured in figure 2.14, where each pinhole is tapered to a diameter of 2 mm. The large diameter for each pinhole is bored through a thickness of 1 cm creating a solid angle of acceptance of 45º. The center positions of each pinhole are listed in the table of figure 2.15(c), where the central pinhole is taken as the origin. The collimators are designed so that the amount of overlap in the solid angles projected onto the reconstruction volume from each pinhole contains the lungs when the magnification is close to 1 [65]. Consequently, data obtained in this way does not maximize the amount of angular coverage, and information for some view angles is missing. Hence, specialized algorithms are required to perform tomographic reconstructions that accurately reflect the distribution of emissions.

(a)

(b)

(c)

Pinhole number	x [cm]	z [cm]
0	0	0
1	-1	-2
2	-2	1
3	2	-1
4	1	2
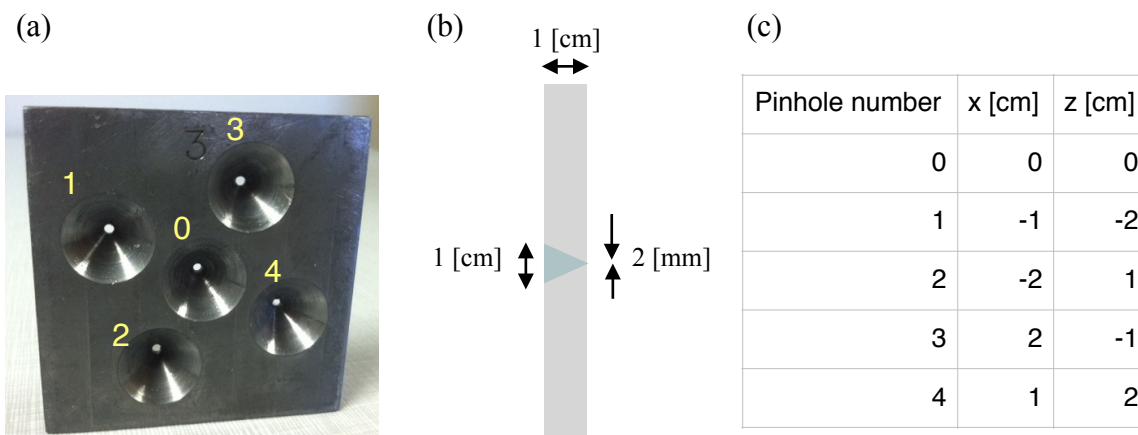
Figure 2.15: The 5 pinhole collimator: (a) photograph of the collimator itself, (b) cross-section of the pinhole, (c) position of the center of each pinhole.

In the past several years, new algorithmic strategies for handling limited angle data have been developed based on the theory of compressed sensing [15, 33] for x-ray CT. These new methods rely on the assumption that some representation of the 3D distribution of activity is sparse. Their use has been demonstrated as an effective means of reconstructing very undersampled CT data. Thus, compressed sensing algorithms in conjunction with multi-pinhole collimators may be effective approaches to reconstructing data obtained from a stationary SPECT system with limited angular coverage.

2.4.1 Preclinical application to lung disease models

Acute lung injury (ALI), occurs in patients with compromised lung function resulting in lungs that cannot properly oxygenate blood [101]. The most serious form of ALI is acute respiratory distress syndrome (ARDS) and is characterized as a rapid progression of hypoxic lung failure [29, 67]. Common causes of ALI include lung infection/sepsis [24], and medical interventions such as complications arising from mechanical ventilation [36], or radiation induced pneumonitis [98]. Treatment often requires mechanical ventilation to deliver high concentrations of oxygen to the lungs. However, prolonged exposure to high levels of oxygen (hyperoxia) can induce oxidative stress in the lungs, cause lung tissue redox changes, and eventually even cell death [24]. In a study of approximately one million patients per year admitted to the ICU 10% of patients requiring invasive ventilation developed ALI [36]. Furthermore, patients who develop ARDS have a 30-40% mortality rate [42, 45, 67]. Recent studies indicate that existing therapies for treating ALI are more effective if detected early [29]. Hence, SPECT biomarkers that could indicate early oxidative stress in the lungs would be a

useful clinical tool for the early detection of ALI, and improve the efficacy of additional interventions.

^{99m}Tc-HMPAO is a radiopharmaceutical used clinically to study cerebral blood flow [71, 97]. It crosses the blood brain-brain barrier and then converts to a form that is trapped within the brain tissue thus serving as a marker of blood flow [4]. Accumulation of HMPAO in lung tissue is based on the same principle: HMPAO is injected in a lipophilic oxidized form and enters lung cells where it is reduced to a hydrophilic form that is trapped in the tissue [6, 71]. The oxidant status of the lung tissue, regulated by glutathione and other factors, controls the extent of trapping so that HMPAO uptake within the lung, and reflects its oxidant status as an indication of disease or injury [6].

In our laboratory, *in vivo* rat models of human lung disease have demonstrated the use of ^{99m}Tc labeled HMPAO as a biomarker for identifying changes in the lung's oxidative state [6, 24, 28]. In one such study, a parallel hole collimator coupled to a NaI gamma camera was used to image the uptake of ^{99m}Tc-HMPAO in the lungs of three groups of Sprague Dawley rats; a control group, a group exposed to 95% oxygen for 24 hours, and a group exposed to 95% oxygen for 48 hours. Dynamic data was obtained for a single view by injecting ^{99m}Tc-HMPAO intravenously and acquiring many frames over a period of 50 minutes. As highlighted in figure 2.16(a), the mean activity of ^{99m}Tc-HMPAO was found in the lungs for each frame, using the boundaries of the lungs as a region of interest (ROI). Likewise, a ROI was placed over the forearm of the animal, and the mean activity in this region was taken as the background activity in the blood. The mean data in both ROIs are shown in figure 2.16(b) over time. Significant changes were

observed in the ratio of the lung to blood steady-state activity in the 24 and 48 hr [99mTc]-HMPAO data compared to the control group.

(a)   (b)
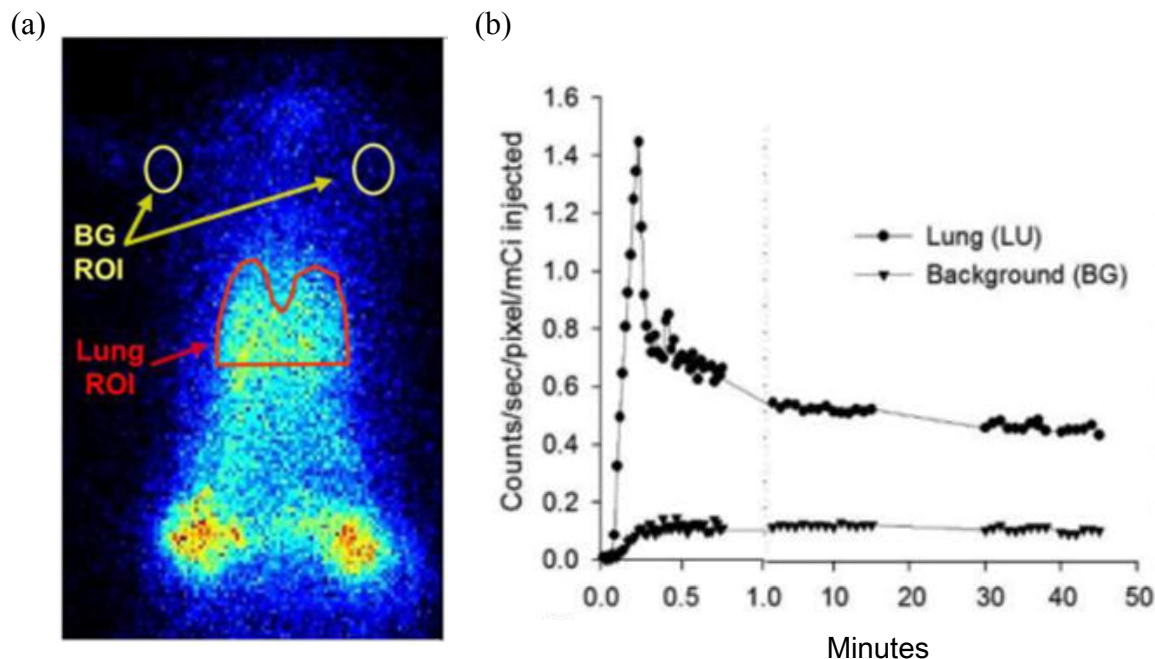
Figure 2.16: (a) Projection obtained from an injection of [99mTc]-HMPAO (Figure 2(A) from [24]), and (b) TAC in the lung and background regions of interest (Figure 4(A) from [24]).

Because [99mTc]-HMPAO may be taken up by other organs and tissues surrounding the lungs, additional imaging is required to obtain the lung boundaries. To this end, an intravenous injection of [99mTc] labeled macroaggregated albumin (MAA) was made after imaging [99mTc]-HMPAO. With a molecule size of 10-100 microns [86], MAA is a clinical pulmonary perfusion marker that gets trapped in the lung microvasculature, and was used in these experiments to identify the lung regions [24]. Due to the close proximity to the lungs, care was taken to exclude the activity originating from the liver. The rat lung anatomy is shown in figure 2.17, depicting frontal view of the lungs. Since data was obtained using a single view angle, pixels positioned at a height on the projection in
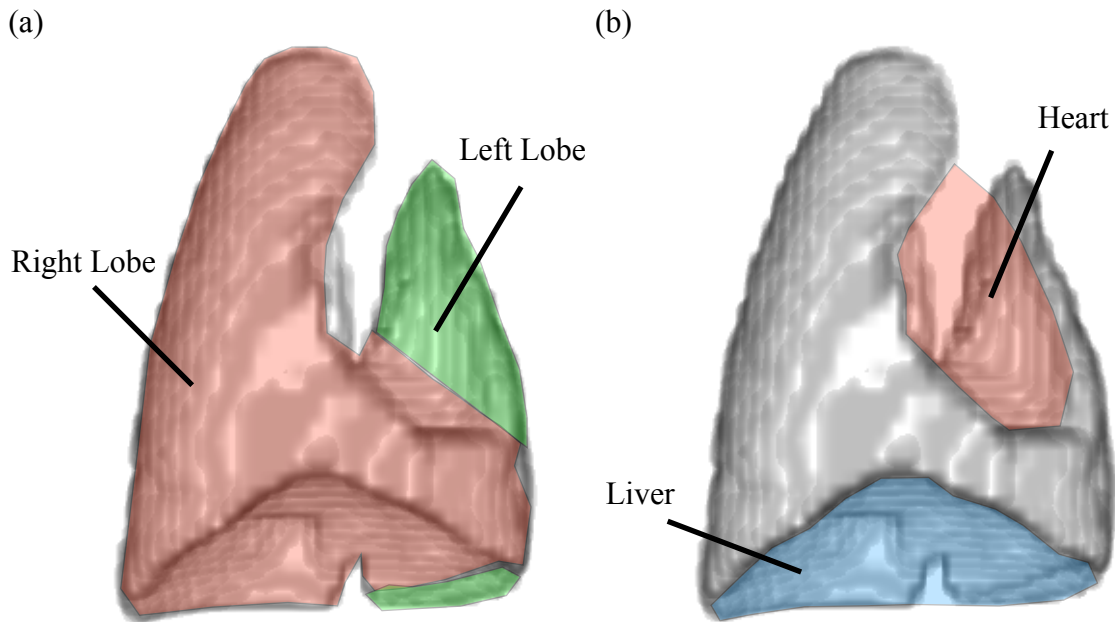
Figure 2.17: Rat lung anatomy: (a) frontal view with the right lobe (red) and the left lobe (green) highlighted, (b) frontal view with location of the heart (red) and liver (blue) highlighted (modified from [82]).

which the liver and lungs overlap could contain counts from either the liver or the lungs. As seen in figure 2.16(a), this results in the bottom of the lungs being clipped from the ROI. Furthermore, the lungs effectively envelop the heart, so that when using only a single view angle, it is not possible to exclude activity originating from the heart in the analysis above. Thus, we propose that an improved estimate of the mean activity in the lungs could be obtained by acquiring data from multiple view angles and performing a 3D tomographic reconstruction on the ^{99m}Tc-HMPAO data. Using ^{99m}Tc-MAA data, a reconstruction of the lung region alone may be obtained and applied as a 3D ROI to the HMPAO reconstructions.

To perform the reconstructions and subsequent analysis proposed here, an efficient implementation of the data model $\mathbf{H}f = g$ must be developed that is both fast

and has a low memory requirement. Chapter 3 describes a matrix-based approach to implementing the forward- and back-projectors that exploits existing matrix-vector multiplication routines and can be performed on a graphics processing unit (GPU).

Chapter 3

Aim 1: Data Model Implementation and Reconstruction

3.1 Introduction

The reconstruction problem in SPECT requires a well-defined mapping from a 3D object space to a set of 2D planar images, represented mathematically by the system matrix **H**. To reduce the computation time in an iterative reconstruction scheme, the most computationally demanding tasks involving **H** may be implemented on a Graphics Processing Unit (GPU), where several thousands of threads may be used simultaneously to perform a computation. This chapter describes an approach to computing and storing a system matrix **H** for use on a GPU, and evaluates the performance of a rotation-based forward- and back-projector.

3.2 Rotation based forward- and back-projector

The first task to forming the system matrix involves calculating the geometric weights using Siddon's method in a ray driven approach. The size of **H** presents a challenge for a practical implementation of the forward- and back-projector. The amount of memory on the current generation of NVIDIA GPUs, for example, ranges from 3-12 GB [73]. For our SPECT system, the size of the detector array is $M = 64^2$ pixels, while the reconstruction volume is $N = 64^3$ voxels. Modeling $V = 64$ view angles, **H** $\in \mathbb{R}^{M \times N \times V}$. Even for this relatively small sized imaging system, **H** would require approximately 275 GB of storage using floating point precision.

An alternative approach would not save **H** to memory at all, and instead calculate the matrix weights using the ray tracing technique for each forward- and back-projection operation. This "on the fly" approach is widely used since it involves less memory, and may be scaled up to much larger reconstruction volumes. As indicated in chapter 2, the ray driven approach is computationally demanding, and calculating matrix weights on the fly may be prohibitive using iterative techniques.

Since the reconstruction methods proposed in chapter 4 may take several thousands of iterations, a compromise between memory and speed has been proposed [2, 19] which takes advantage of existing fast matrix multiplication routines. The memory footprint of **H** can be minimized in a number of ways. We begin by observing that projections from different view angles may be obtained using multiple cameras placed at different angles around the subject, or using a single gamma camera by rotating the subject and acquiring the data at each angle. Digitally, we may approach the forward- and back-projection operations in the same way. Using multiple cameras represents an entirely different set of rays for each view angle, whereas the rotation approach requires only a single set of rays. To reduce the storage size, **H** need only be calculated and stored for a single view angle, or $\bar{\mathbf{H}}$. A complete forward-projection may be obtained by iteratively rotating the image volume about the z-axis and multiplying by $\bar{\mathbf{H}}$. The pseudo-code for the rotation-based forward-projection is shown in algorithm 3.1, where each

$$
\begin{aligned}
&1: \quad \textbf{for } \phi < nAngles \\
&2: \quad\quad g_\phi = \bar{\mathbf{H}}\hat{f} \\
&3: \quad\quad \text{rotate } \hat{f} \text{ in the } + \phi \text{ direction} \\
&4: \quad \textbf{end}
\end{aligned}
$$

Algorithm 3.1: Pseudo code for the rotation-based forward-projection.

projection g_ϕ represents an entire 64^2 pixel vector corresponding to the ϕ view angle. The amount of storage for $\bar{\mathbf{H}}$ is then approximately 4.3 GB, as in the example above, which is still quite large.

The amount of memory used to store $\mathbf{H}$ can be further reduced by only storing the non-zero elements of $\bar{\mathbf{H}}$. Due to the pinhole geometry for SPECT, only a limited subset of voxels are actually in the line of sight of a given detector element. Each row of $\bar{\mathbf{H}}$ will therefore contain mostly zeros, and may be stored using a sparse matrix format. The format for this dissertation is Compressed Sparse Row (CSR) format, due to the fact that the number of rows of $\bar{\mathbf{H}}$ is generally much less than the number of columns. CSR format stores the matrix using three 1D arrays; a *value*, *column*, and *pointer* array (by convention). The *value* array contains the floating point value of each $\bar{h}_{ij}$ element in row major order. The *column* array contains the column location corresponding to each element in the *value* array. The *pointer* array is a running total of elements for each row in the system matrix. Figure 3.1 illustrates CSR format for a simple sparse matrix using zero-based indexing.

$$
\begin{bmatrix}
a & 0 & b & 0 \\
0 & c & 0 & 0 \\
0 & 0 & d & e \\
0 & 0 & f & 0
\end{bmatrix}
\longrightarrow
\begin{array}{ccc}
\textit{value} & \textit{column} & \textit{pointer} \\
\begin{bmatrix} a \\ b \\ c \\ d \\ e \\ f \end{bmatrix} &
\begin{bmatrix} 0 \\ 2 \\ 1 \\ 2 \\ 3 \\ 2 \end{bmatrix} &
\begin{bmatrix} 0 \\ 2 \\ 3 \\ 5 \\ 6 \end{bmatrix}
\end{array}
$$
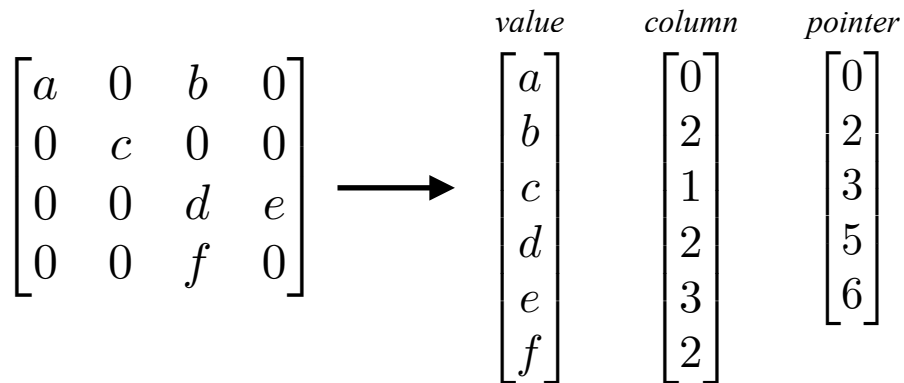
Figure 3.1: A simple matrix converted into CSR format.

The overall size of $\bar{\mathbf{H}}$ depends on the imaging geometry and the number of pinholes, which changes the ray density, the size of the pinhole mask in figure 2.14(b), and hence the number of nonzero entries in the matrix. However, by storing the matrix in CSR format for a single view, the matrix used for reconstructing the experimental data in Chapter 6 is only 17.6 MB, when $N = 64^2$ and $M = 64^3$, and is computationally more efficient.

The back-projection operation may be carried out in a similar manner as the forward-projector with two major differences; the matrix multiplication requires $\bar{\mathbf{H}}^T$, and the rotation operation must be performed in the opposite direction of the forward-projection. Since the transpose will contain more columns than rows, $\bar{\mathbf{H}}^T$ will be stored in Compressed Sparse Column (CSC) format.

Like CSR format, CSC format stores the matrix in three 1D arrays; *value*, *row*, and *pointer*. The *value* array in CSC format contains the elements $\bar{h}_{ij}$ as they appear in column major order. The *row* array contains the row location corresponding to the elements in *value*, and the *pointer* array is the running total of elements in each column. This is illustrated in figure 3.2 for the same simple matrix as figure 3.1. Notice, the matrix in CSR format is the transpose in CSC, and vice versa. In this way, $\bar{\mathbf{H}}^T$ may be

Figure 3.2: A simple matrix converted into CSC format.

formed in CSC format by taking the *column* array from $\bar{\mathbf{H}}$ and making it the *row* array for $\bar{\mathbf{H}}^T$, and using the same *pointer* array from $\bar{\mathbf{H}}$.

The back-projection may also be carried out by digitally rotating the reconstruction volume. However, a rotation of the reconstruction volume by ϕ is equivalent to a rotation of the detector by $-\phi$. When performing the back-projection, the frame of reference is with respect to the detector, hence we must rotate in the opposite direction of the forward-projection, where the frame of reference is with respect to the reconstruction volume. The rotation based back-projector is summarized in algorithm 3.2.

$$
\begin{aligned}
&1: \quad \textbf{for } \phi < nAngles \\
&2: \qquad \text{rotate } \hat{f} \text{ in the } -\phi \text{ direction} \\
&3: \qquad \hat{f} = \hat{f} + \bar{\mathbf{H}}^T g_\phi \\
&4: \quad \textbf{end}
\end{aligned}
$$

Algorithm 3.2: Pseudo-code for the rotation-based back-projection.

3.2.1 The rotation method

As indicated in the section 3.2, the volume rotation is an integral part of implementing the forward- and back-projection operators, and can even impact the convergence behavior of the CP algorithm. To begin, the affine rotation transformation [1, 38] is used on *xy*-coordinates of the digital volume,

$$
\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix}, \tag{3.1}
$$

where each *xy*-plane in the volume is rotated about the *z*-axis. Our implementation only accepts angles ϕ between -90º and 90º. For equally spaced $\theta \notin$ [-90º, 90], θ is converted to ϕ that has identical projections, in magnitude, on the *x* and *y* axes according to

$$\phi = |\theta| - q \cdot \pi/2 \ , \tag{3.2}$$

where *q* is the quadrant of the original angle θ. In order that the transformed coordinates fall on a well defined angular grid, ϕ in equation 3.2 is rounded according to

$$\phi = sign(\theta)\lfloor \phi/\Delta\theta - 1/2 \rfloor \Delta\theta \ . \tag{3.3}$$

Figure 3.3 illustrates how rotations for all angles are obtained using ϕ. The angle θ is recovered by transposing the rotated coordinates generated from equations 3.1 and 3.3, or
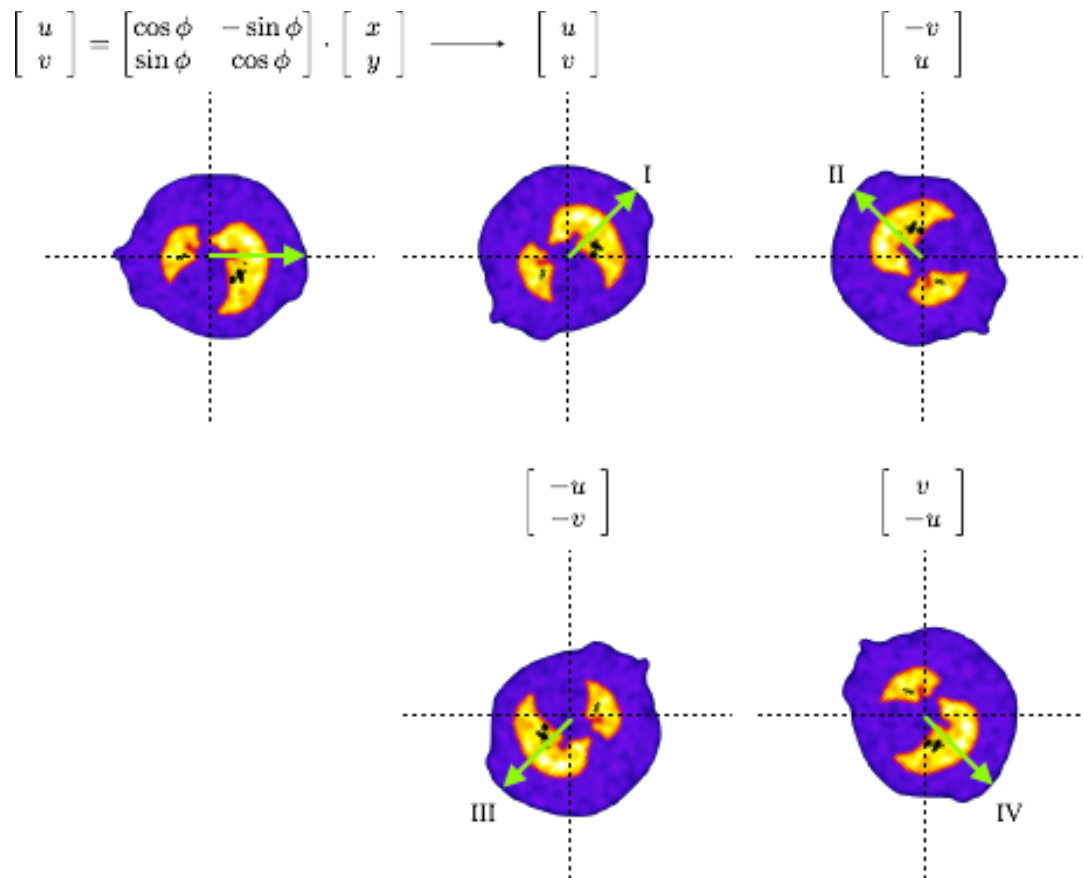
Figure 3.3: Illustration of the rotation method.

reflecting the rotated coordinates about their respective axis, depending on the desired quadrant and the direction of rotation. In this way, angles in each quadrant are formed on the same grid of angular displacements, ensuring the back-projection is the transpose of the forward-projection.

Finally, using the grayscale values corresponding to the transformed coordinates, bilinear interpolation [1, 38] is used to assign grayscale values to the voxels that fall on the coordinate grid, thus preserving the total grayscale energy (the sum of all grayscale intensities) in the transformed plane. Although bilinear interpolation introduces a blur in the rotated volume, it does not affect the sum over rays traced through the rotated volume.

3.3 Software overview

For datasets consisting of multiple acquisitions over time, software that is capable of processing multiple datasets simultaneously reduces processing time, particularly if there are a large number of acquisitions. The computational task may be distributed among multiple GPUs to speed up the processing of multiple datasets. To this end, a distributed C++ program was implemented which utilizes the parallel processing libraries provided by MPI (Message Passing Interface) [69] in order to compute $\bar{\mathbf{H}}$ and to pass $\bar{\mathbf{H}}$ to multiple GPUs.

In MPI, multiple threads may be instantiated representing independent and parallel threads performing identical tasks. For example, a system consisting of two GPUs may initialize two MPI threads, given by their "rank". The rank 0 process communicates with GPU 0 and the rank 1 process communicates with GPU 1. Since

multiple processes are initialized, the calculation of the system matrix may be distributed across the MPI threads as illustrated in figure 3.4.

For our implementation, using two MPI threads as an example, two halves of the system matrix $\bar{\mathbf{H}}$ may be computed simultaneously, where the rank 0 process computes the first half of $\bar{\mathbf{H}}$ and rank 1 computes the second half $\bar{\mathbf{H}}$. The task of computing each half of $\bar{\mathbf{H}}$ may be accelerated using local parallelism. Each MPI process may call their own local threads using OpenMP [77], to divvy up the computation of their respective

Figure 3.4: Schematic of the software implementation.

rows. Once the two halves are calculated they are combined and the full $\bar{\mathbf{H}}$ is distributed to both MPI processes, where it is converted into CSR format, normalized, then distributed to the GPUs. Additionally, the system matrix may be saved for subsequent reconstructions, so as not to compute $\bar{\mathbf{H}}$ each time the program is invoked.

The current implementation performs the reconstruction on the GPUs using NVIDA® CUDA^{TM} (Compute Unified Device Architecture) [75], where the forward- and back-projection operations are carried out using an established matrix multiplication routine and a custom kernel to perform the rotation operation. This approach takes advantage of the cuSparse library's fast matrix-vector multiplication method [74], while also leveraging several thousands of parallel threads to perform the volume rotation. Since the reconstruction method may require several thousand iterations, the reconstruction is performed entirely on the GPU, thus avoiding the need to pass arrays back and forth between the CPU and GPU at each iteration.

3.4 GPU system matrix performance

Performance of the forward- and back-projection operation was evaluated using the execution time of the rotation method and the cuSparse matrix-vector multiplication method. Simple experiments were performed using two NVIDIA K-20c Tesla GPUs[1] and using a 2.3 GHz Intel i5 CPU (4 cores) [49] for comparison. To test the matrix multiplication routine, a $64^2 \times 64^3$ CSR system matrix with 2,466,080 non-zero matrix elements was used. This particular matrix models our SPECT system's geometry and was used to perform the reconstructions presented in chapter 5.

[1] The two GPUs used for this dissertation were provided by Dr. Taly Gilat-Schmidt, installed on an Exxact [31] linux server.

To begin, the rotation method was tested using 3, 6, 15, and 60 view angles. For each view angle, 100 rotations were performed, the mean and standard deviation were calculated, and plotted in figure 3.5. On average, one rotation of the entire 64^3 volume takes approximately 0.164 ± 0.1 ms, where performance is degraded slightly for an odd number of view angles. As might be expected from the description of the rotation method, the exact time of execution was found to depend on the quadrant and the angle's displacements from ϕ. For ϕ angles close to 0°, the transformed coordinates do not differ much from 0°, and therefore the (x, y) coordinates are close to the transformed coordinates (u, v) as laid out in memory. As ϕ gets closer to 90°, the (x, y) and (u, v) coordinates are no longer in the same or consecutive blocks of memory, leading to ~0.5
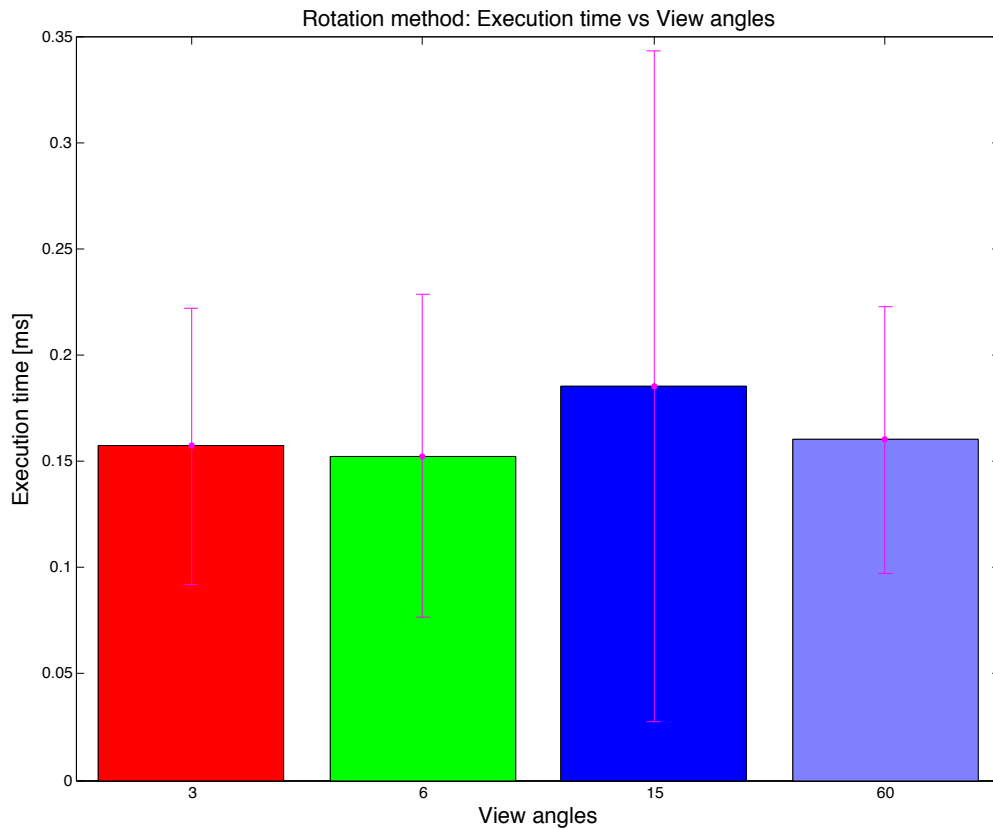
Figure 3.5: Average execution time for a rotation using 3, 6, 15, and 60 equally spaced view angles.

ms longer run times. This is demonstrated in figure 3.6 for 60 equally spaced view angles. Using our method, there was no noticeable penalty incurred for angles requiring reflections and transpositions. For comparison, the same rotation method was implemented in MATLAB [66] and run on a CPU using a single thread, where the method takes 0.20 ± 0.04 ms to rotate a single plane of the 64^3 digital volume.
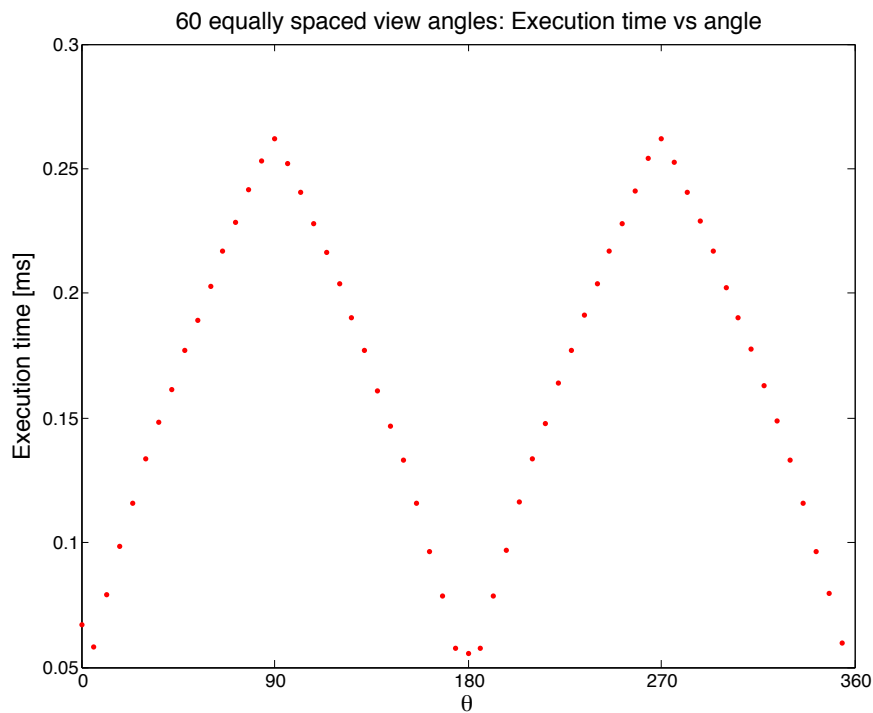
Figure 3.6: Execution time vs angle for 60 equally spaced view angles.

Likewise, the performance of the cuSparse floating point matrix-vector multiplication method (run on a GPU) was evaluated and compared to MATLAB's builtin sparse matrix multiplication method. The system matrix was loaded onto the GPU, and then the matrix and a simple digital volume were multiplied 100 times. On the GPU, the matrix-vector multiplication run time was 0.23 ± 0.015 ms. Similarly, on the CPU the sparse matrix multiplication run time was 6.0 ± 0.85 ms.

3.5 Conclusion

A rotation-based, sparse-matrix approach to the forward- and back-projection operators was developed and will be used in the subsequent chapters. It allows for CT reconstructions on a single GPU, when large memory resources are not available. Utilizing a GPU and established matrix multiplication methods, our implementation demonstrates improvement in speed of execution compared to an equivalent CPU implementation. The cuSparse matrix-vector multiplication is about 26% faster than MATLAB's built in sparse matrix-vector routines. Additionally, whole 64^3 digital volumes may be rotated on a GPU at roughly the same time that it takes to rotate a single plane on a CPU. For reconstructions requiring several thousand iterations, the combination of matrix multiplication and rotation on a GPU presents a significant reduction in the overall reconstruction time in an iterative scheme. Furthermore, using distributed parallelism, the reconstructions could be carried out on multiple GPUs, allowing multiple reconstructions to be performed simultaneously. Alternatively, reconstructions could be performed on multiple computers, each with their own GPU(s).

Chapter 4

Aim 1: 3D Reconstruction for Pinhole SPECT

4.1 Overview

Mathematically reconstructing a distribution of radioactivity emissions from a set

of projection data requires an inversion procedure for the Radon transform. This chapter

begins by providing an overview of these methods which fall into two categories; direct

"Fourier based" methods, and iterative methods. The two iterative methods used in this

dissertation, Chambolle-Pock (CP) and the Maximum Likelihood Expectation

Maximization (MLEM) method, are then presented. Additionally, the role of noise and

concepts of compressed sensing is discussed in the context of computed tomography.

4.1.1 Direct methods

If f is an unknown object and g is the projection data obtained from the object,

the mathematical problem in SPECT, and in CT in general, is to invert the Radon

transform of $f(x, y)$. The Radon transform of $f(x, y)$ is given by

$$g(t, \theta) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y)\, \delta(x\, cos\theta + y\, sin\theta - t)\, dx\, dy \tag{4.1}$$

where a parallel ray model is assumed and a single 2D slice of the reconstruction is

considered to simplify the discussion. An analytical approach to inverting equation 4.1,

i.e. finding f from a set of projections g, is based on the Central Slice theorem [9, 53],

which relates the Radon transform to the Fourier transform $G_\theta(\omega)$,

$$G_\theta(\omega) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y)\, e^{-2\pi i \omega(x cos\theta + y sin\theta)}\, dx\, dy \tag{4.2}$$

This is the basis for the widely used filtered back-projection [53] method given by

$$\hat{f}(x,y) = \int_0^{2\pi} g(t,\theta) * h(t) \Big|_{t=x\,cos\theta+y\,sin\theta} d\theta \quad , \qquad (4.3)$$

where the form of the filter $h(t)$ can be represented by $|\omega|$ in the frequency domain.

Computationally, methods based on filtered back-projection, such as Feldkamp's [32], are straightforward and very fast. 1D FFT (Fast Fourier Transform) [26] algorithms for filtering come standard in most software packages and the back-projection operation is only required once.

The usefulness of Fourier based methods are limited by the Shannon-Nyquist sampling theorem [76, 87]. In figure 4.1 a simple digital phantom was reconstructed using filtered back-projection with 4, 8, 16, and 180 view angles. The reconstructions

(a)

(b)

(c)

(d)
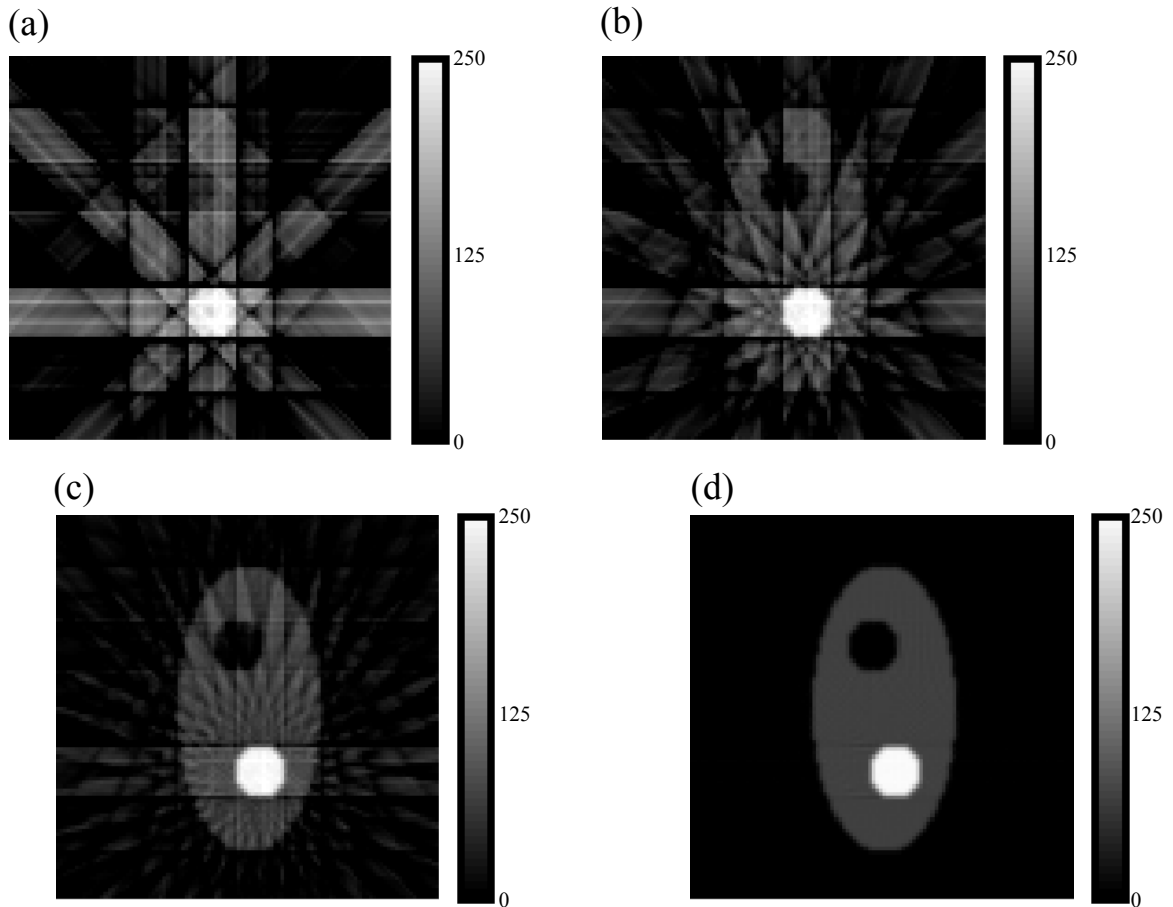
Figure 4.1: Reconstruction of a simple phantom using filtered back-projection using (a) 4 views, (b) 8 views, (c) 16 views, and (d) 180 views.

using 4, 8, and 16 view angles contain streak artifacts where the back-projector assigns activity outside the boundary of the phantom. This illustrates that filtered back-projection is not well suited for stationary micro-SPECT systems that are constrained to only a few view angles.

The use of high pass filters $h(t)$ required by equation 4.3 also limits the usefulness of filtered back-projection methods in SPECT. For micro-SPECT a large number of projections are required to reconstruct small scale features in $f(x, y)$ and to minimize streak artifacts in the reconstruction. Furthermore, SPECT data has a larger variance in the signal compared to x-ray CT. Unless the filter is well designed, $h(t)$ will give too much weight to high frequency content in the Fourier domain, resulting in noisy reconstructions.

4.2 Iterative reconstruction

Standard iterative techniques, such as ART (Algorithmic Reconstruction Technique)[38, 56] and MLEM (Maximum Likelihood Expectation Maximization)[56, 88], treat the problem of obtaining reconstructions from a set of projections as a linear inversion problem, $\mathbf{H}f = g$, where again f is the unknown object, g is the projection data, and $\mathbf{H}$ is the imaging or system operator. Although there are many iterative techniques, the general approach to the tomographic problem is similar; begin with an initial guess of the solution, compare the resulting forward-projection of the guess to the actual data, update the guess to better satisfy some functional that incorporates a measure of how well the "guessed" data matches the measured data, and repeat the process until certain convergence criteria are met. These methods of successively better

approximations are generally aimed at reducing streak artifacts, as seen in figure 4.1, while also incorporating models that reflect the physics of the imaging system and prior knowledge of the solution into the reconstruction.

4.2.1(a) Maximum likelihood expectation maximization (MLEM)

In SPECT, MLEM is an iterative method that seeks the maximum likelihood estimate of the distribution of photon emissions by incorporating the statistics of the emissions and detections as part of the reconstruction model. The objective of MLEM is to find a solution whose mean distribution of emissions maximizes the likelihood of producing the observed data [9, 68, 88]. The following description is summarized from Shepp and Vardi [88].

The statistical interpretation of the detection process begins by considering the probability an emission from within the reconstruction volume will be detected by one of the detector elements. Suppose the reconstruction volume is a 3D ensemble of N independent voxels, where each voxel contains some amount of a radioisotope, and the detector is a 2D array of M independent detectors, as in figure 4.2. For a single emission
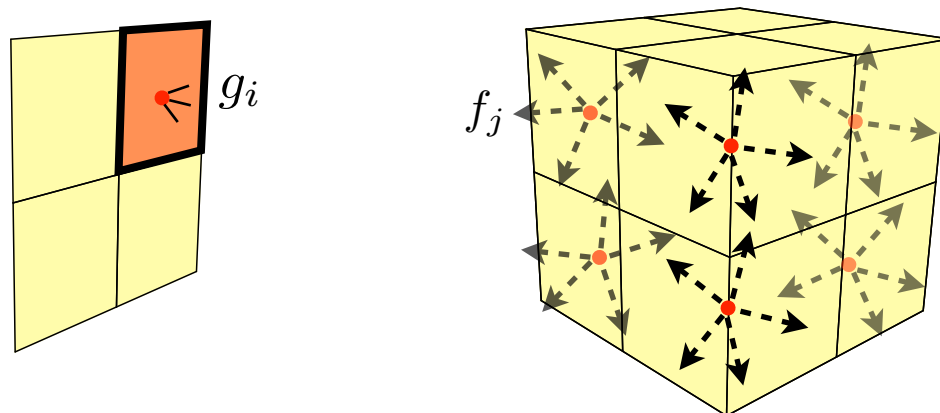
Figure 4.2: The imaging system modeled by an array of M independent detectors, on the left, and an ensemble of N independent voxels, on the right, each containing an unknown number of radioactive emitters.

originating from the jth voxel, the probability of being detected by the ith detector element depends on the projection model and the random process that models the physics of the emission. If we let the probability of an emission in the jth voxel be P_j, then the probability of detecting that emission in the ith detector, P_i, may be stated as

$$P_i = P_j(h_{ij}|\mu) , \tag{4.4}$$

where h_{ij} models the image forming collimator and μ is a physical property (the mean activity) of the radioactive emitter.

A projection g is the accumulation of detections over the acquisition process. Since emissions may originate from anywhere within the reconstruction, the probability of obtaining g_i counts in the ith detector element is the sum over the probabilities from all voxels given by

$$P(g_i) = \sum_{j=0}^{N-1} P_j(h_{ij}|\mu) . \tag{4.5}$$

If we assume the counts in each detector element, g_i, are independent of one another, the probability of obtaining the vector of all independent detector elements, g, is the product of all M independent probabilities $P(g_i)$, or the joint distribution of all $P(g_i)$, and is given by the likelihood function $L(g)$, where

$$L(g) = P(g_0) \cdot P(g_1) \cdot \ldots P(g_{M-1}) = \prod_{i=0}^{M-1} \sum_{j=0}^{N-1} P_j(h_{ij}|\mu) . \tag{4.6}$$

Events within the jth voxel are generated by the spontaneous emission of a gamma photon from an unstable nucleus. The rate of emissions from voxel j of a radioactive isotope over time, or the isotope's *activity*, a_j, is directly proportional to the concentration of isotopes contained in that voxel, n_j, so that

$$a_j = \frac{dn_j}{dt} = -\lambda \cdot n_j . \tag{4.7}$$

The constant of proportionality, λ, is the decay rate constant for the specified radionuclide. Using the solution to the differential equation in 4.7, and solving for a_j shows the activity decays exponentially over time,

$$a_j(t) = a_j^o \, e^{-\lambda t}$$ (4.8)

where a_j^o is the initial activity within the jth voxel. It can be shown that the probability of an emission event with activity modeled by equation 4.8 is Poisson distributed [53, 68, 88], i.e.,

$$Pr(k|\bar{\mu}) = \bar{\mu}^k \frac{e^{-\bar{\mu}}}{k!}$$ (4.9)

where k is the number of emissions and $\bar{\mu}$ is the mean activity.

For a given time, t, during the acquisition, either a detector element registers an event or it does not, as in a Bernoulli process. Statistically then, a projection is the sum of many Bernoulli events generated by a Poisson process, and is Poisson distributed [9, 68].

Incorporating the probability in equation 4.9 with the projection model $\mathbf{H}$, we can formulate the likelihood of measuring a complete set of projection data, g, from an emission object f using a SPECT system consisting of M detector elements and N object voxels as

$$L(g|f) = \prod_{i=0}^{M-1} \sum_{j=0}^{N-1} (h_{ij} f_j)^{a_{ij}} \left(\frac{e^{-h_{ij} f_j}}{a_{ij}!} \right)$$ (4.10)

The objective of MLEM is then to maximize $L(g|f)$ and is traditionally accomplished using the method proposed by Shepp and Vardi [88], which performs an expectation step on the log likelihood prior to maximizing the resultant. From [88], if we let $\hat{f}_j^{(n)}$ be the

estimate after the nth iteration in the jth voxel, the Shepp and Vardi approach yields a fixed point iteration procedure given by

$$\hat{f}_j^{(n+1)} = \frac{\hat{f}_j^{(n)}}{\sum\limits_{i=0}^{M-1} h_{ji}} \sum_{i=0}^{M-1} h_{ji} \frac{g_i}{\sum\limits_{k=0}^{N-1} h_{ik}\hat{f}_k^{(n)}} , \qquad (4.11)$$

and is detailed in Appendix B for completeness.

4.2.1(b) MLEM implementation and reconstructions

According to equation 4.11, in one MLEM iteration, the algorithm formulates a multiplicative update to the current estimate of the emission distribution, $\hat{f}^{(n)}$ as illustrated in figure 4.3. The update begins by comparing the data, g, to the forward-projection of $\hat{f}^{(n)}$, known as pseudo data, by taking their ratio, r. Back-projecting r

$$\textit{initialize all } \hat{f}_j^{(n)} \textit{ to one}$$

$$\hat{f}(n) \longrightarrow g_i/(\mathbf{H}\hat{f}^{(n)})_i$$

$$\hat{f}_j^{(n)} \cdot w_j \longleftarrow \mathbf{H}^T r = w$$
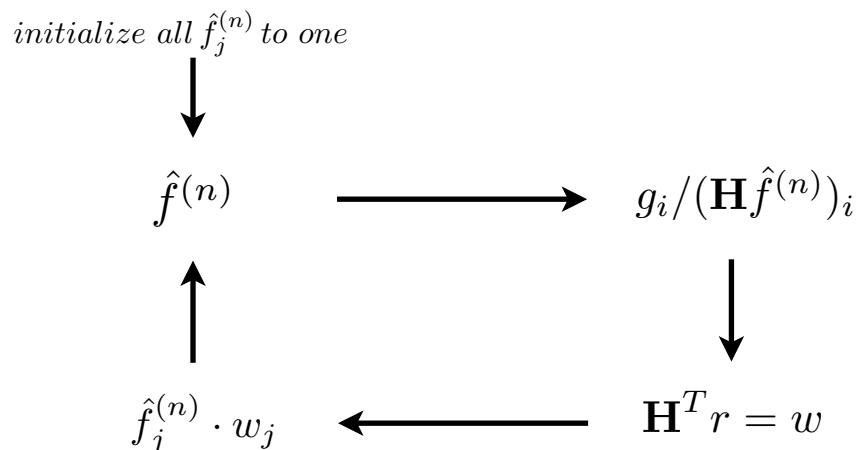
Figure 4.3: A schematic of the MLEM iterative procedure (modified from [56]).

creates a distribution of corrections, w, in $\mathbb{R}^N$ that is used to re-weight the current estimate. Finally, the iteration concludes by dividing by the back-projection of h_{ij}. Since the elements h_{ij} represent probabilities, back-projecting h_{ij} will result in a distribution of 1's in regions of the reconstruction volume that are visible from all view angles. Regions

that are not visible to all view angles will have values less than one. In practice, rather than back-projecting h_{ij} for each iteration, the operation may be replaced by masking the rows of the system matrix **H** which are only visible from all view angles.

Since MLEM requires one forward- and one back-projection operation for each iteration, it is computationally more intensive than filtered back-projection. On the other hand, reasonable reconstructions are achieved using far fewer views than that required for filtered back-projection [53]. For example, In figure 4.4 MLEM reconstructions where performed on the simple phantom from figure 4.1. Digital projections were taken at 16 equally spaced views around the phantom by applying the forward-projection on the digital volume. Reconstructions are shown using 1, 10, 20, and 30 iterations. Comparing
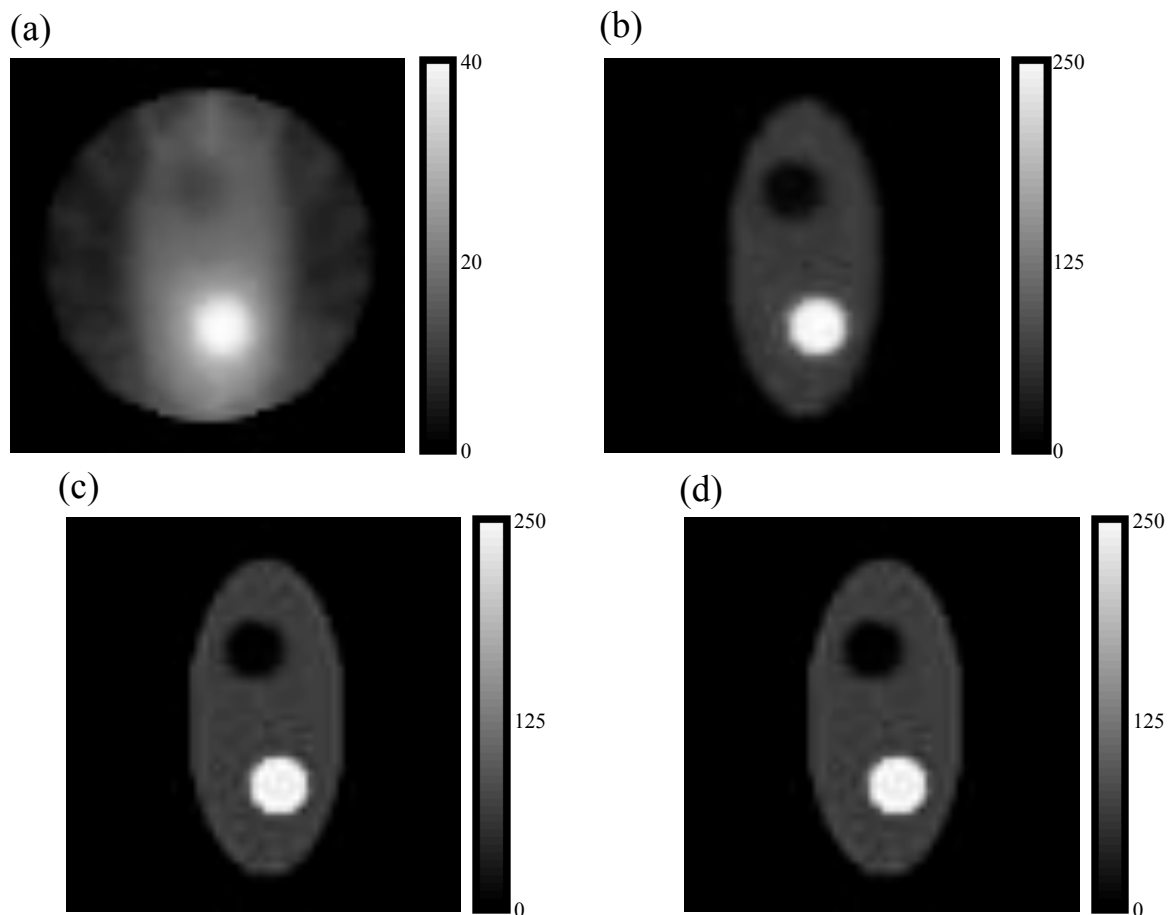
Figure 4.4: MLEM reconstructions of a simple phantom using 16 views; (a) 1 iteration, (b) 10 iterations, (c) 20 iterations, and (d) 30 iterations.

figure 4.4(d) obtained using MLEM to figure 4.1(c) using filtered back-projection,

MLEM appears to reduce streak artifacts in the reconstructions but suffers sharpness.

4.2.2 The SPECT objective function, noise, and constrained optimization

Pseudo-data $\mathbf{H}\hat{f}$ generated from a good estimate of the emission distribution

should agree with the measured data g. This is known as *data fidelity*. One common

measure of data fidelity is the Euclidean distance given by

$$D_{EU}(g, \mathbf{H}f) = \sqrt{\sum_{i=0}^{M-1} [\mathbf{H}f - g]_i^2} \quad . \qquad (4.12)$$

If the elements of the vectors $\mathbf{H}\hat{f}$ and g are generated by a Gaussian random process, as

is the case in x-ray CT, then the expected distance between the vectors is given by the

Euclidean norm.

Likewise, if the elements of the vectors $\mathbf{H}\hat{f}$ and g are generated by a Poisson

random process, as is the case in SPECT, then the expected value of the difference

between the two vectors is given by the Kullback-Leibler divergence, or D_{KL}, where

$$D_{KL}(g, \mathbf{H}f) = \sum_{i=0}^{M-1} \left[\mathbf{H}f - g + g \, log\left(\frac{g}{\mathbf{H}f}\right)\right]_i \quad \text{[9]} \qquad (4.13)$$

The KL distance between the pseudo-data and the actual data should be close to 0 if the

estimate $\hat{f}$ accurately reflects the true distribution of emissions, f. Then the SPECT

inversion procedure may be written as the optimization problem,

$$\hat{f} = \underset{f:\mathbf{H}f=g}{\arg\min} \{D_{KL}(g, \mathbf{H}f)\} , \qquad (4.14)$$

where the optimal solution $\hat{f}$ is selected from the set $\{f \mid \mathbf{H}f = g\}$ to minimize D_{KL}.

Minimizing the Kullback-Leibler distance using MLEM has been used in SPECT reconstruction, and shown to be effective at reducing streak artifacts in the reconstruction of under-sampled data [88]. There are drawbacks to using MLEM, however. Since data obtained in the lab contains noise due to photon scatter and the acquisition process, performing minimization solely on a data metric without imposing any constraints on the object can result in solutions with excellent data fidelity, but at each iteration of MLEM the likelihood of obtaining the noisy data is increased. The result will be noisy solutions that do not accurately reflect the actual mean distribution of emissions. To avoid this noise amplification, MLEM is typically stopped after 30 to 50 iterations [68, 88], limiting the algorithm's ability to resolve the boundaries in the reconstruction for very under-sampled data.

One approach to controlling noise amplification is to introduce a penalty term into the optimization problem. The modified problem has a solution that minimizes the data error D_{KL} subject to a constraint operationally defined by $P(f)$, i.e.,

$$\hat{f} = \underset{f:\mathbf{H}f=g}{\arg\min} \left\{ D_{KL}(g, \mathbf{H}f) + \mu P(f) \right\}, \tag{4.15}$$

where μ controls the "tradeoff" between good data fidelity and noise amplification [9]. When used to control noise, the functional $P(f)$ is constructed to penalize solutions that are not smooth.

4.2.3 Compressed sensing

An effective approach to the constrained optimization problem in equation 4.15 is to assume that the imaged object or some representation of the object is "sparse" (or contains many zeros). The idea of using "sparseness" to promote smooth estimates of the object is illustrated in figure 4.5 for a noisy signal [64]. Suppose we have a signal (figure 4.5(a)) that is degraded by noise (figure 4.5(b)). An approach to reconstructing 4.5(a) from 4.5(b) is to assume the solution is piecewise smooth, using methods which choose solutions with minimal variance between the signal "spikes." The solution to this simple de-noising problem is composed mostly of zeros.

(a)   (b)
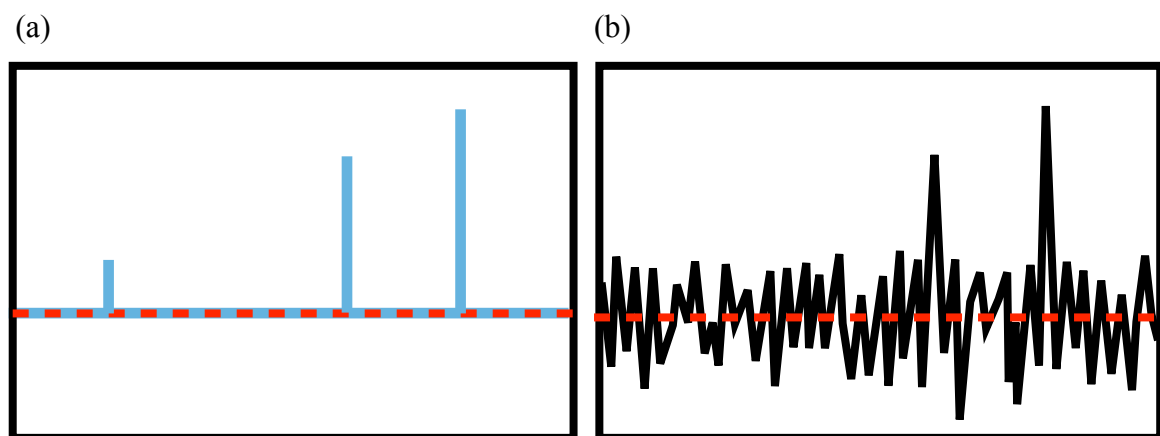
Figure 4.5: (a) A sparse signal consisting of "spikes," and (b) a sparse signal with noise (adapted from figure 2 in [64]).

Methods that exploit sparsity in the system fall under the umbrella of compressed sensing algorithms. Originally developed by Candes, Tao, and Donoho [15], compressed sensing is a general mathematical theory for finding solutions to linear systems that are underdetermined. The theory states that a linear system may be recovered using far less

data than required by the Shannon-Nyquist requirement, provided some basis of the

solution is sparse.

To present an overview of compressed sensing, consider the following linear

system,

$$Ax = b \tag{4.16}$$

where $A \in \mathbb{R}^{M \times N}$, $x \in \mathbb{R}^N$ and $b \in \mathbb{R}^M$. If $M < N$ and if x is to be estimated within

some degree of accuracy, A must sufficiently sample x. In the context of CT, this means

the system matrix must generate a large number of view angles from the object.

Otherwise, if $M << N$, which is generally the case in computed tomography, there are

too many degrees of freedom to meaningfully estimate x and an exact solution is not

possible.

If x contains many zeros and k is the number of nonzero values in x, such that

$1 \leq k \leq N$, then x is said to be "k-sparse" [99]. According to the theory of compressed

sensing, the system matrix A need only sample x at least $2k$ times [52, 99, 33]. If

$M < 2k$, there is a vector s_{2k} containing $2k$ nonzero elements in the null space of A, i.e.

$A s_{2k} = \mathbf{0}$, which may be written as the difference between two vectors $s_k^{(1)}$ and $s_k^{(2)}$ that

contain k nonzero entries. Hence,

$$A s_{2k} = A \left(s_k^{(1)} - s_k^{(2)} \right) = \mathbf{0}$$
$$A s_k^{(1)} = A s_k^{(2)} , \tag{4.17}$$

and the $M \times N$ matrix gives the same data b for two different k nonzero vectors.

Therefore, there is no way to uniquely recover a k-sparse signal if $M < 2k$.

This implies that for an underdetermined linear system that is k-sparse, if A has at

least $2k$ linearly independent columns, then a solution may be recovered using a

procedure that 1) determines the independent columns of A, and 2) uses some optimization routine to minimize the number of non-zero elements in x. In other words, the solution x to the system is the sparsest x that satisfies $Ax = b$ where the number of non-zeros in x is the metric for sparsity, i.e.

$$x = \arg\min_{x:Ax=b} \|x\|_0 \ . \tag{4.18}$$

This measure is known as the zero-norm

$$\|x\|_0 = \lim_{p \to 0} \|x\|_p = \sum_{i=0}^{N-1} |x|_i^0 \ . \tag{4.19}$$

Unfortunately, practical optimization methods do not exist for 0-norm minimization. Instead, the 1-norm, is typically used as a surrogate measure of sparsity, i.e. where $\|x\|_1$ is the sum of all elements of x.

$$\|x\|_1 = \sum_{i=0}^{N-1} |x|_i \ , \tag{4.20}$$

Figure 4.6 depicts the 0-norm unit ball on the left, and the 1-norm unit ball on the right, where the highlighted regions contain all points that make up vectors with respective norm less than or equal to 1.

(a)
$$\|x\|_0 = \sum_{i=0}^{N-1} |x|_i^0 \leq 1$$

(b)
$$\|x\|_1 = \sum_{i=0}^{N-1} |x_i| \leq 1$$
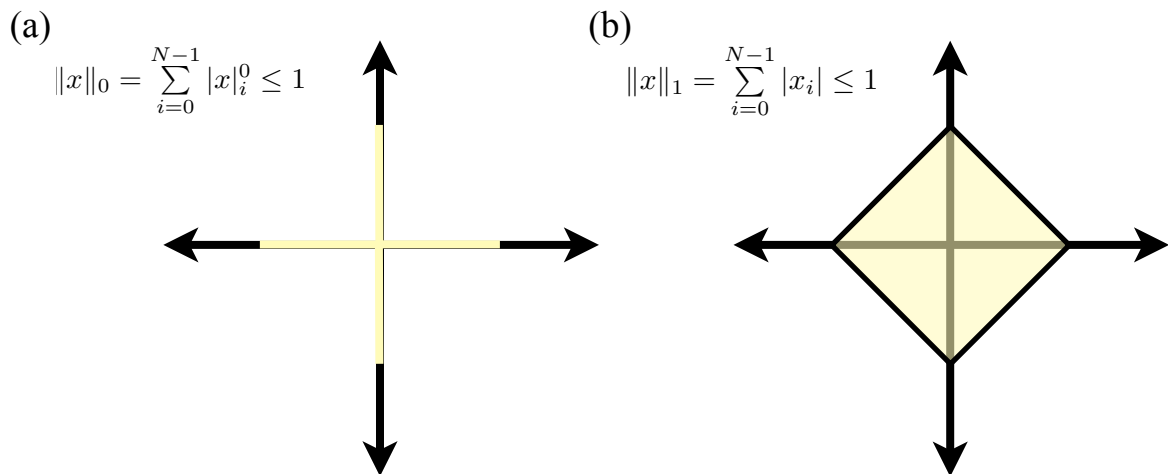
Figure 4.6: The vector space representation of (a) the 0-norm unit ball, and (b) the 1-norm unit ball. Regions highlighted in yellow represent the set of all vectors that are less than or equal to 1 in norm.

Intuitively, if the the vector x contains many zeros its 1-norm should be small.
The 1-norm is a useful approximation to the 0-norm because the boundary of the 1-norm
unit ball is the convex hull of the 0-norm unit ball. Many well established convex
optimization routines may therefore be employed for 1-norm minimization [72], where 1-
norm minimization is known as "basis-pursuit," and is illustrated in figure 4.7. The blue
line represents the set of all possible solutions to a linear system $Ax = b$ in 2D. The 1-
norm is used to find the sparsest solution that satisfies the linear system. In this approach
the 1-norm ball is expanded about the origin of a 2D vector space where the intersection
of the 1-norm ball and the solution set is used to pick out the vector with the greatest
number of zeros, which minimizes the magnitude of x, and satisfies $Ax = b$. The 1-norm
solution in this simple illustration happens to also be the 0-norm solution.
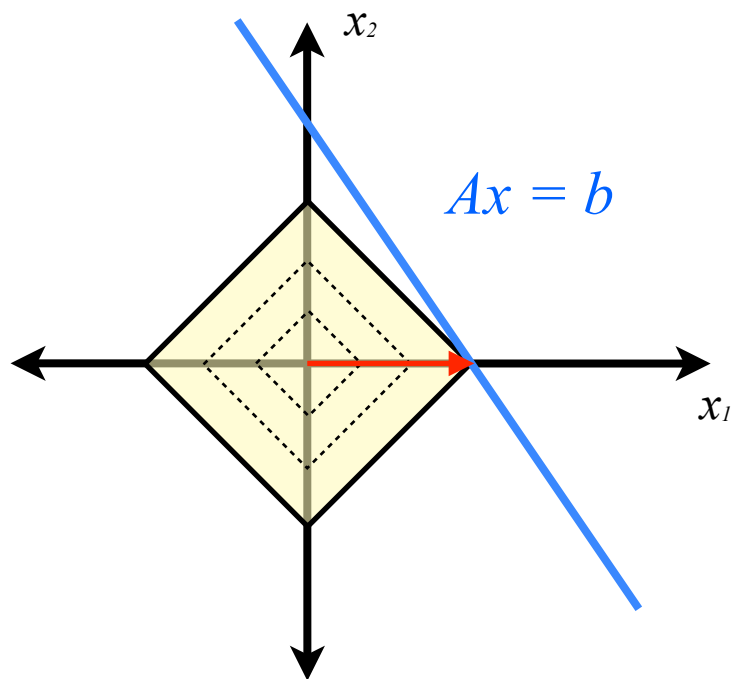
Figure 4.7: An illustration of basis-pursuit, where the 1-norm selects a solution from the
    set.

The solution to many emission tomography problems $\mathbf{H}f = g$ are not generally sparse. There may, however, be an expansion of some basis of f, given by x such that

$$x = \mathbf{\Psi}f ,$$

(4.21)

where the solution x to equation 4.18 is known to be sparse and $\mathbf{\Psi}$ is the corresponding "sparsifying" operation. The minimization procedure described above may therefore be carried out in the basis defined by $\mathbf{\Psi}$. The new minimization problem may be written as

$$f = \arg\min_{f:\mathbf{H}f=g} \|\mathbf{\Psi}f\|_1 .$$

(4.22)

For systems with large dimensionality, Candes, et al. [15] demonstrated that the solution obtained through 1-norm minimization in the $\mathbf{\Psi}$-basis is equivalent to the 0-norm solution with high probability provided the system is under-sampled in a way that is "incoherent" to the null-space of the $\mathbf{\Psi}$-basis [15, 52, 33, 99]. If we let A represent the full system matrix and $\mathbf{\Phi}$ represent a matrix consisting of an independent subset of A, the coherence between $\mathbf{\Phi}$ and $\mathbf{\Psi}$ is given by the maximum correlation between the two,

$$u(\mathbf{\Phi}, \mathbf{\Psi}) = \max\left\{\langle\mathbf{\Phi}, \mathbf{\Psi}\rangle\right\} ,$$

(4.23)

where u is a measure of the degree to which the system may be undersampled. It has been shown that incoherent sampling, when u is small, can be accomplished by randomly sampling the system [15]. Furthermore, random sampling is a much more efficient strategy for determining an independent subset of columns in A than using an exhaustive search [99].

For computed tomography, randomly sampling the system matrix A implies taking projections from random view angles, a process for which a conventional CT system is generally not designed. Moreover, stationary SPECT systems have gamma

cameras fixed at regularly spaced angular intervals. Further work has extended the concepts of compressed sensing to systems with noise [41], signals that are not terribly sparse [70, 104], and systems that employ undersampling on regular intervals [37, 51]. Compressed sensing may therefore be applied to data collected from conventional imaging systems. Furthermore, more data may be collected per view angle in SPECT using multi-pinholes. Conventional micro-SPECT systems, such as FastSPECT-II [34] and U-SPECT II [10, 100], employ multi-pinhole collimators in order to obtain more projection data per view, provide a larger field-of-view (FOV), and reduce the signal-to-noise ratio (SNR) [65, 100]. This motivates our hypothesis that using multi-pinhole data in combination with sparsity promoting reconstruction methods may work well for SPECT data with very few views.

An implementation of equation 4.15 for obtaining the optimal solution requires the determination of a sparse basis. There has been much success in the past five years using the gradient magnitude of the reconstructed image as a sparse basis in x-ray CT [91, 92, 93]. A three component gradient of the image is formed by taking the first-order image derivative in the x, y, and z directions, to create three edge volumes. The magnitude image is then formed by taking the square root of the sum of squares of the three edge volumes. To demonstrate, figure 4.8(a) depicts the Shepp-Logan phantom [89] with its gradient magnitude image (right). In this case, the gradient magnitude image is zero everywhere except where there are edge features in the original image. If the 1-norm is used as the measure of sparsity, the 1-norm of the gradient magnitude of f, known as "total-variation" (TV), is given by,

$$\mathbf{TV} = \| \, |\nabla f| \, \|_1 = \sum_{i=0}^{N-1} |\nabla f|_i \, . \tag{4.24}$$

(a) f (b) $|\nabla f|$
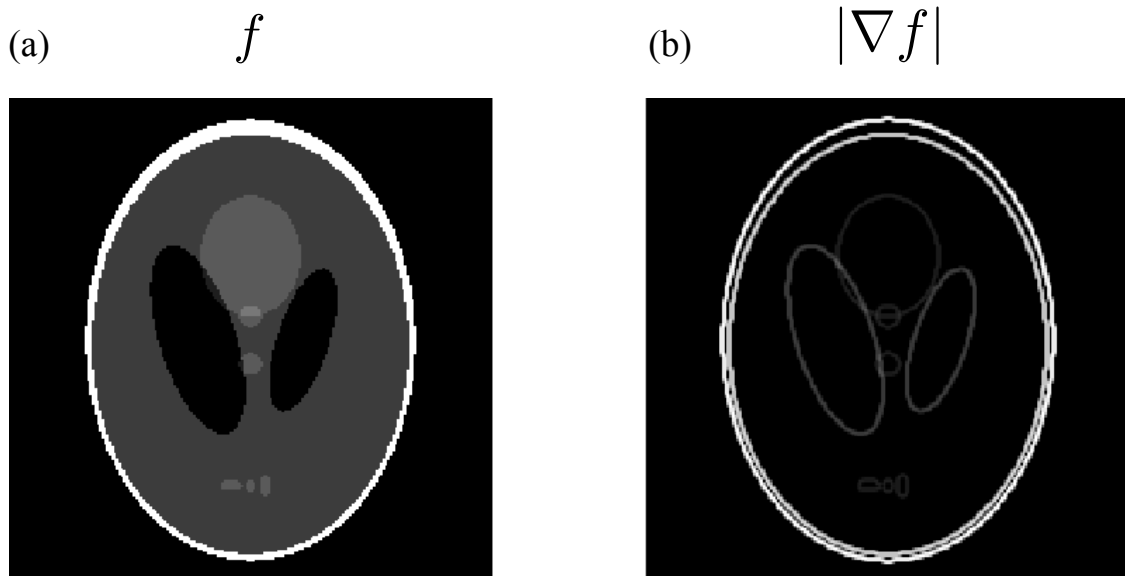
Figure 4.8: (a) The Shepp-Logan phantom [89], and (b) a sparse representation of the Shepp-Logan phantom using its gradient magnitude image.

In this case, we expect the TV to be small because the gradient magnitude contains mostly zeros. Although there are other sparse bases that could be used, total-variation is the most widely used in CT applications, and will be the basis used in the SPECT problem of this dissertation.

When the total-variation is used in the constrained optimization problem of equation 4.15, the SPECT reconstruction problem is to determine $\hat{f}$ such that

$$\hat{f} = \underset{f:\mathbf{H}f=g}{\arg\min} \left\{ D_{KL}(g, \mathbf{H}f) + \mu \| |\nabla f| \|_1 \right\}. \tag{4.25}$$

This incorporates the assumption that the gradient magnitude of f is sparse. Since minimizing the total-variation promotes sparseness in the gradient magnitude of the solution, solving equation 4.25 has the potential to work well with under-sampled data. The task remains to identify an appropriate optimization routine which minimizes both the Kullback-Leibler data fidelity measure and the total-variation in the solution.

4.2.4(a) Chambolle-Pock (CP) algorithm and generalized descent

One approach to solving equation 4.25 for $\hat{f}$ is to use an iterative descent method. Although descent methods vary in their implementation, successive approximations are generated in a similar way. Consider a general problem in which we seek an optimal solution x^* that minimizes Kx. Then, for a given x_n , an approximation that is closer to x^* can be obtained as

$$x_{n+1} = x_n + \sigma K x_n \text{ ,} \tag{4.26}$$

using a linear update Kx_n scaled by an appropriate step size, σ. In gradient descent the linear operator, K, is the gradient of some scalar potential function f

$$x_{n+1} = x_n + \sigma\, \partial f(x_n) \text{ ,} \tag{4.27}$$

where the objective is to seek an x^* such that $\partial f(x)$ is minimized. The gradient of the potential f is used to generate the search direction and the step size σ is used to control how far to move in the direction of the gradient, as illustrated in figure 4.9, where each contour represents a level curve of f. If σ is small, many steps are required to reach x^*. If σ is large fewer steps are required, but the method may overshoot x^*. An overview of
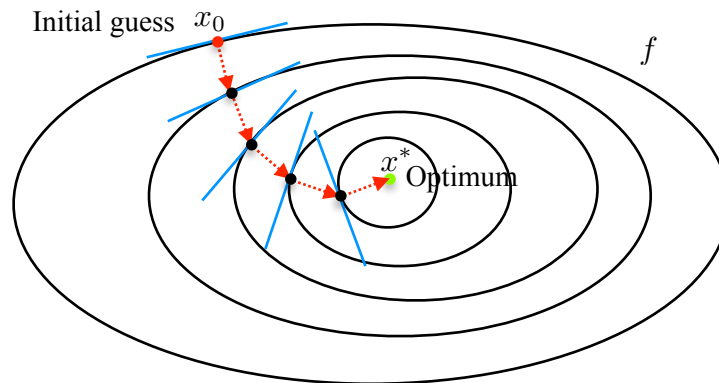
Figure 4.9: Optimization using gradient descent in 2D. The contours represent a level set of the scalar potential function f for a given x .

some common descent methods used in computed tomography, including gradient descent and conjugate gradient, may be found in [9].

The TV-norm in our proposed constrained optimization problem, equation 4.25, precludes the direct use of the gradient descent method, since the TV-norm is not smooth in general, as evident in figure 4.8(b). The descent method proposed by Chambolle and Pock is an alternative descent approach which may be used to optimize such objective functions. Provided the functionals are convex, and satisfy the first order convexity condition, the CP method makes no requirements on the smoothness of the objective function, and in this sense the method is a more general descent approach than gradient descent, in equation 4.27.

From calculus, the max or min of a smooth function may be found by solving for points where the slope is zero and checking the concavity at those points. In general, a differential function f is convex if there are "supporting hyperplanes" such that the inequality in figure 4.10 holds for all x. Geometrically, the hyperplane in figure 4.10 is the first order Taylor approximation to the function about the point a, i.e. the function $f(x)$ lies above the hyperplane tangent to f for all points a. Functions meeting this

$$f(a) + (\partial f(a))^T (x - a) \leq f(x)$$
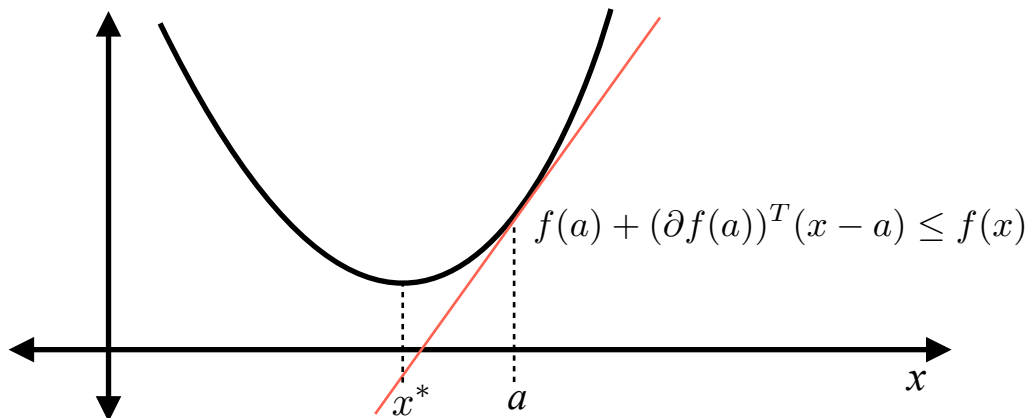
Figure 4.10: An illustration of the first-order convexity condition.

criteria, known as the first-order convexity condition [12], are said to have a global

minimizer, x^* where the supporting hyperplane's slope is zero. The optimization can then

be formulated as solving

$$\partial f(x) = 0 \,.$$
(4.28)

If the convex function is not differentiable everywhere the first-order convexity

condition may be rewritten in terms of the function's subdifferential, defined by [12, 84]

$$\partial f(a) = \left\{ z \in \mathbb{R}^N : f(a) + z^T(x - a) \leq f(x), \ \forall\, x \in \mathbb{R}^N \right\},$$
(4.29)

and illustrated in figure 4.11. Convex functions with "kinks," take $f(x) = |x|$ for

example, may have a set of supporting hyperplanes with subgradients z for a given x.

The convex function has a global optimizer x^* if the zero vector is contained in the

function's subdifferential,

$$\mathbf{0} \in \partial f(x^*) \,.$$
(4.30)

In the example depicted in figure 4.11, $|x|$ has a set of supporting hyperplanes $z \in [-1,\ 1]$

at $x = 0$. Since 0 is contained in the subgradient at $x = 0$, $x = 0$ is a global optimizer of

$|x|$.
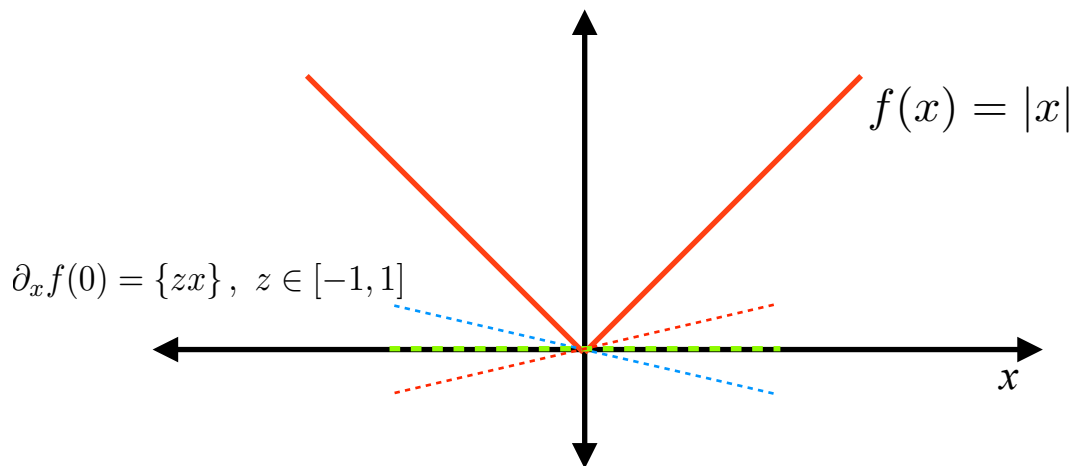
Figure 4.11: An illustration of the subdifferential at $x = 0$ for $f(x) = |x|$.

For a convex function that is not smooth there is an iterative descent procedure

analogous to the gradient descent approach known as a proximal point iteration,

$$x_{n+1} = (\mathbf{I} + \sigma\,\partial f)^{-1}\, x_n = \text{prox}_\sigma[f](x_n) \qquad (4.31)$$

where ∂f is the subdifferential of an objective function f, σ is the step size, and the

resolvent [63] of the inversion procedure $(I + \sigma\partial f)^{-1}$ is defined as the proximal

mapping [11, 83]. A more detailed discussion of using the proximal mapping for iterative

optimization is presented in appendix C for completeness.

The iterative procedure proposed by Chambolle and Pock simultaneously solves

two problems, a primal and dual problem, using proximal mappings to perform the

optimization. Consider minimizing the objective function of the form,

$$\min_{x\in X}\left\{F(Kx) + G(x)\right\}, \qquad (4.32)$$

where $x \in X$ and $y \in Y$, and the matrix K is the forward mapping from X into Y. If the

functions F and G are convex, a dual objective function may be constructed by taking the

negative convex conjugate of the primal problem in equation 4.32 to obtain,

$$\max_{y\in Y}\left\{-F^*(y) - G^*(-K^T y)\right\}, \qquad (4.33)$$

where F^* denotes the convex conjugate of F [17].

The convex conjugate is taken by performing the Legendre transform [12] on the

functions F and G, defined for a function F by

$$F^*(y) = \max_x\left\{y^T x - F(x)\right\}^{.} \qquad (4.34)$$

The Legendre transform is illustrated in figure 4.12 for a 1D function. Consider the set of

all affine functions such that $y^T x + b \leq F(x)$. In figure 4.12, a unique line is chosen by
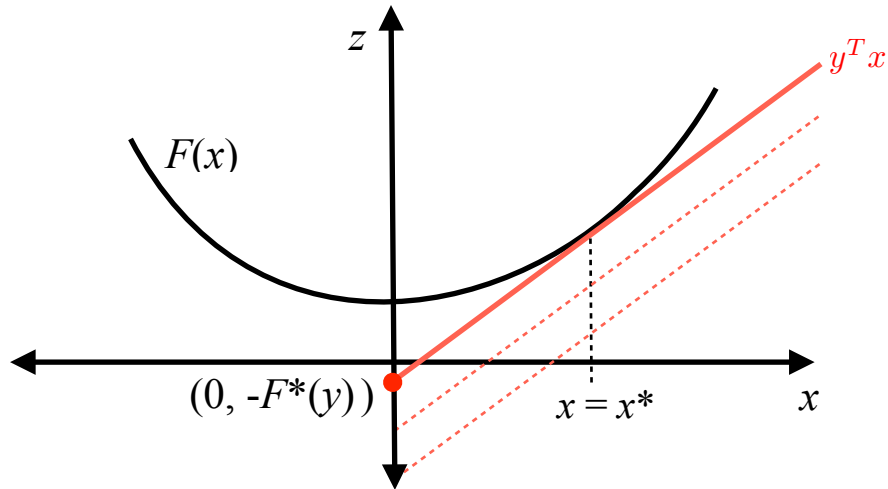
Figure 4.12: An illustration of the Legendre transform of a 1D convex function $F(x)$ (adapted from figure 3.8 in [12]).

selecting an intercept b, then $y^T x + b \leq F(x)$ approaches the convex function $F(x)$ from below. The intercept is maximal when the affine function is a supporting hyperplane of the function $F(x)$, where the two functions intersect at x^*. That is, $y^T x^* + b^* = F(x^*)$, which implies $y \in \partial F(x^*)$. Solving for b^* we obtain

$$b^* = F(x^*) - y^T x^*$$
$$-F^*(y) = \max_x \left\{ F(x) + y^T x \right\}$$

(4.35)

By convention, the convex function in terms of the transformed variable y is defined as $b^* = -F^*(y)$.

Equation 4.32 and 4.33 imply that the optimization problem given in equation 4.25 may be solved by minimizing the objective function, or alternatively by maximizing the negative convex conjugate of the objective function. Thus the optimal solution to the SPECT problem may be obtained by solving the original primal problem or its corresponding dual problem. Conceptually, the primal and dual problem form a convex min-max, or saddle-point, optimization problem as depicted in figure 4.13.

Primal problem

$$\min_{x \in X} \{F(Kx) + G(x)\}$$

Dual problem

$$\max_{y \in Y} \{-F^*(y) - G^*(-K^T y)\}$$
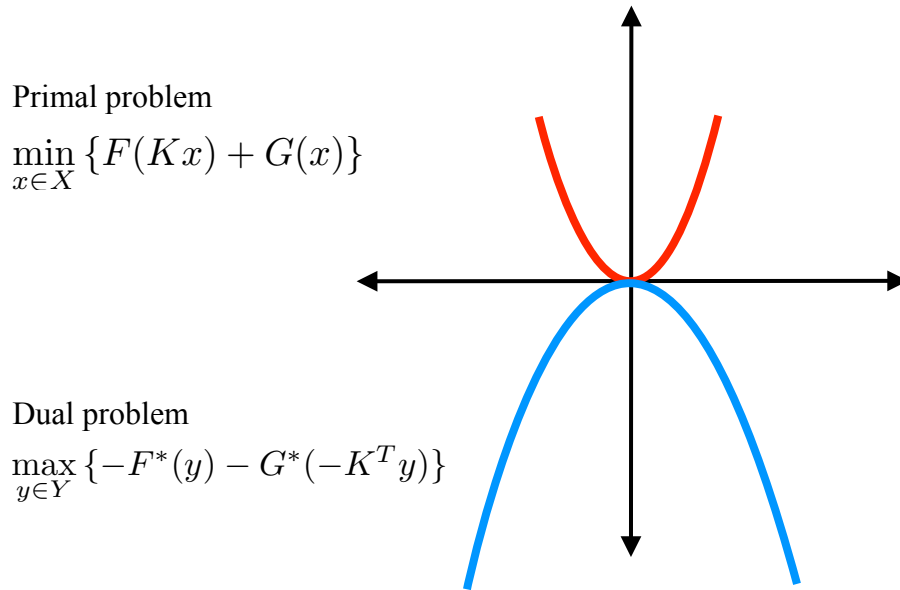
Figure 4.13: 2D visualization of a convex primal-dual problem, where the dual problem (bottom curve) is the negative convex conjugate of the primal problem (top curve).

The saddle-point optimum to the primal and dual objective functions, x^* and y^*, occurs when the "gap" between the primal and dual potentials is zero [17, 91, 104]. The CP method, therefore, seeks a solution that simultaneously optimizes the primal and dual problems. In an iterative scheme, the difference between the primal and dual objectives, known as the "primal-dual gap" [17] may be used as a non-heuristic [92] means of assessing how close an estimate is to this optimum.

Chambolle and Pock proposed an iterative descent algorithm that utilizes the duality formed by the Legendre transform in algorithm 4.1. The saddle-point optimum, x^* and y^*, implies [17, 84],

$$Kx^* \in \partial F^*(y^*)$$
$$-K^T y^* \in \partial G(x^*) \ .$$

(4.36)

What is notable about equation 4.36 is the dual formulation is quite appropriate to CT problems, as the mappings performed by K and K^T naturally extend to the forward- and

back-projection operations. Additionally, an iterative proximal procedure may be derived

from equation 4.36 that is similar to the proximal point iteration in equation 4.31. This

leads to the general CP method shown in algorithm 4.1. Lines 4 and 5 of the algorithm

are the proximal iterations for the dual formulation, and are used to generate the descent

direction on line 6. A detailed derivation of the proximal mappings for equation 4.36 is

presented in appendix C. In algorithm 4.1, τ and σ are the step sizes of the primal and

dual problems, respectively, and L is the largest singular value of the forward matrix of

the linear system.

$$
\begin{aligned}
&1: \quad L = \|K\|_2; \quad \tau = 1/L; \quad \sigma = 1/L; \quad \theta = 1; \quad n = 0 \\
&2: \quad \text{initialize vectors } \mathrm{x}_0, \ \mathrm{y}_0, \ \text{and } \bar{\mathrm{x}}_0 \ \text{to } \mathbf{0} \\
&3: \quad \textbf{repeat} \\
&4: \qquad \mathrm{y}_{n+1} \leftarrow \mathrm{prox}_\sigma[F^*](\mathrm{y}_n + \sigma K \bar{\mathrm{x}}_n) \\
&5: \qquad \mathrm{x}_{n+1} \leftarrow \mathrm{prox}_\tau[G](\mathrm{x}_n - \tau K^T \mathrm{y}_{n+1}) \\
&6: \qquad \bar{\mathrm{x}}_{n+1} \leftarrow \mathrm{x}_{n+1} + \theta(\mathrm{x}_{n+1} - \mathrm{x}_n) \\
&7: \qquad n \leftarrow n + 1 \\
&8: \quad \textbf{until convergence}
\end{aligned}
$$

Algorithm 4.1: The Chambolle-Pock iterative procedure.

Because the CP iterative procedure utilizes proximal mappings to generate the

descent search direction, it may be used to optimize objective functions with non-smooth

constraints, such as total-variation. In fact, the method has been extended to TV-penalized

reconstructions in x-ray CT [91, 93] and to 2D SPECT [92, 104].

4.2.4(b) The CP algorithm for 3D SPECT

We now formulate the procedure for solving the 3D SPECT problem using the CP

method. First we take a side-by-side comparison between the primal objective in (4.32),

$\min_{x \in X} \{F(Kx) + G(x)\}$, and the constrained objective function for SPECT in (4.25),

$\hat{f} = \arg\min_{f:\mathbf{H}f=g} \{D_{KL}(g, \mathbf{H}f) + \mu\| |\nabla f| \|_1\}$. This leads to an iterative procedure based on

Algorithm 4.1 for determining the optimal solution to the SPECT problem using the CP

algorithm with TV-KL minimization. Details can be found in [92] and Appendix C, with

a brief description here.

The key task is to tailor F and G to the SPECT problem, then find their

corresponding convex conjugates. One way to simplify the analysis is to let F in the

primal problem $\min_{x \in X} \{F(Kx) + G(x)\}$ be the entire objective function,

$$F = F_1(\mathbf{H}f) + F_2(\nabla f) = D_{KL}(\mathbf{H}f, g) + \mu\| |\nabla f| \|_1 \quad (4.37)$$

and to let the function G be zero. The system matrix $K = \begin{pmatrix} \mathbf{H} \\ \nabla \end{pmatrix}$ then performs the

forward-projection $\mathbf{H}$ multiplication, and the gradient image operation, ∇. The convex

conjugate of F is formed by taking the Legendre transform of F, which is defined by

$$F^*(z) = \max_x \{x^T z - F(x)\} \quad (4.38)$$

for the variable x and the transformed variable z. The resulting convex conjugate is

$$F^*(y', \vec{z}') = F_1^*(y') + F_2^*(\vec{z}') = \sum_i [-g \ln(\mathbf{1} - y')]_i + \delta_\mu(|z'|) \quad (4.39)$$

where $\vec{z}'$ is a three component volume formed by the image gradient operation, δ_μ is an

indicator function, and $y' \in \mathbb{R}^M$.

Then the proximal mapping of F^*, and in particular $prox_\sigma[F_1^*]$ for the SPECT

problem is given by

$$prox_\sigma[F_1^*](y) = \tfrac{1}{2}\left((\mathbf{1} + y) \pm \sqrt{(\mathbf{1} - y)^2 + 4\sigma g}\right) \quad (4.40)$$

as described in Appendix C. The CP method for TV-penalized D$_{KL}$ minimization follows

from equation 4.40 and the proximal mappings, and is given in algorithm 4.2.

1 : INPUT : data g, data-error tolerance ϵ, primal-dual scaling s
2 : INPUT : system matrix $\mathbf{H}$, smoothing parameter ν, TV-penalty λ
3 : initialize f_0, $\bar{f}_0$, y_0, $\vec{z}_0$ to zero
4 : $L = \|\mathbf{H}, \nu\nabla\|_2$; $\tau = s/L$; $\sigma = 1/(sL)$; $\theta = 1$; $n = 0$;
5 : **repeat**
6 : $\quad y_n' = \frac{1}{2}(1 + y_n + \sigma\mathbf{H}\bar{f}_n - \sqrt{(y_n + \sigma\mathbf{H}\bar{f}_n - 1)^2 + 4\sigma g})$
7 : $\quad y_{n+1} = \max(\|y_n'\|_2 - \sigma\epsilon, 0)\frac{y_n'}{\|y_n'\|_2}$
8 : $\quad \vec{z}_n' = \vec{z}_n + \sigma\nu\nabla\bar{f}_n$
9 : $\quad \vec{z}_{n+1} = \vec{z}_n'((\lambda/\nu)/\max(\lambda/\nu, |\vec{z}_n'|))$
10 : $\quad f_{n+1} = f_n - \tau(\mathbf{H}^T y_{n+1} + \nu\nabla^T \vec{z}_{n+1})$
11 : $\quad \bar{f}_{n+1} = f_{n+1} + \theta(f_{n+1} - f_n)$
12 : $\quad n = n + 1$
13 : **until convergence**
14 : OUTPUT : f_N
15 : OUTPUT : y_N, $\vec{z}_N$

Algorithm 4.2: The CP method for TV penalized D$_{KL}$ minimization.

4.2.4(c) The CP design parameters

There are a number of parameters that may affect the convergence and

performance of the Chambolle-Pock method. In particular, an appropriate level of

smoothing for the TV-penalized D$_{KL}$ minimization problem must be determined, while

the relative scaling between the primal and dual problems can affect our choice in step

sizes.

Since the penalty term, μ, is used to control noise in the reconstructions, we may

expect with more noise in the data, the larger μ must be to effectively reduce the amount

of variation in the solution. If too much weight is given to the penalty term, the algorithm

may require many iterations to fully realize solutions with good data fidelity. Likewise, if

μ is too small we may quickly arrive at solutions with good data fidelity that give too much weight to noise, as in the case of running MLEM for too many iterations.

Additionally, the scaling on the magnitude of D_{KL} and the magnitude of the TV in our objective function may be quite different. It is proposed in [91] to introduce an additional parameter ν that effectively converts the "units" of the gradient operation to the units of the forward-projection. A standard conversion may be defined as the ratio of the largest singular value of $\mathbf{H}$ over the largest singular value of the ∇ matrix,

$$\nu = \nu_{crit} = \|\mathbf{H}\|_2 / \|\nabla\|_2 \tag{4.41}$$

where the gradient is performed using finite difference matrices. The constrained optimization procedure may then be rewritten as

$$\hat{f} = \arg\min_{f}\left\{D_{KL}(\mathbf{H}f, g) + \tfrac{\lambda}{\nu}\| \, |\nu\nabla f| \, \|_1\right\} \tag{4.42}$$

where ν is kept fixed at ν_{crit} and λ may be adjusted for a specific application.

Like many descent algorithms, the choice of step sizes σ and τ can affect the number of iterations required to reach an acceptable solution or can cause the routine to become unstable. If σ and τ are not scaled well, the optimization routine can become unstable, or an inordinate number of iterations may be required for convergence. Chambolle and Pock explicitly demonstrate convergence of algorithm 4.1 when σ and τ are given by,

$$\sigma\tau < 1/L^2 \tag{4.43}$$

and L is the largest singular value of the system matrix K [17]. Conceptually, σ and τ are related to one another through a property of the Legendre transform, and numerically scaled according to the eigenvalues of the system matrix $\mathbf{H}$. In the case of the Legendre

transform, the step sizes are inversely proportional to one another, and are scaled at most

by the largest singular value of the system matrix. The Fourier transform exhibits a

similar property [9], and is illustrated in figure 4.14. Namely, a very localized function in

x is transformed to function that is dispersed in frequency u.
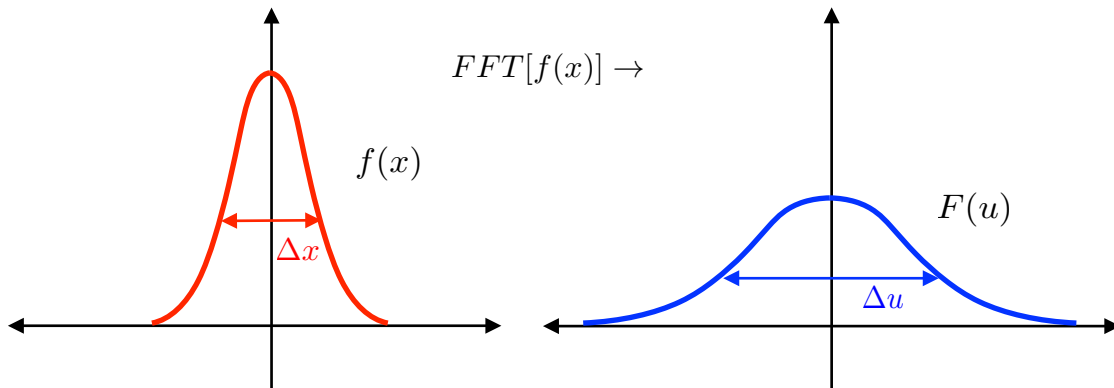
Figure 4.14: A narrow function transforms to a broad function in the Fourier domain (modified from figure 3.10 in [9]).

Appropriate step sizes σ and τ reflect how localized or dispersed the primal and

dual problem are relative to one another, as in figure 4.13 and 4.15. Although $F(x)$ and

$-F^*(y)$ are plotted in the same plane in figure 4.15, in general x and the transformed

coordinates y are not in the same dimension. A scaling parameter s is proposed to

determine appropriate values of σ and τ such that

$$\sigma = 1/(sL)$$
$$\tau = s/L$$

(4.44)

since equation 4.43 tells us nothing about their relative sizes. Practical implementation

issues will be discussed in chapter 5.2, including the choice of s, λ, and stopping criteria.
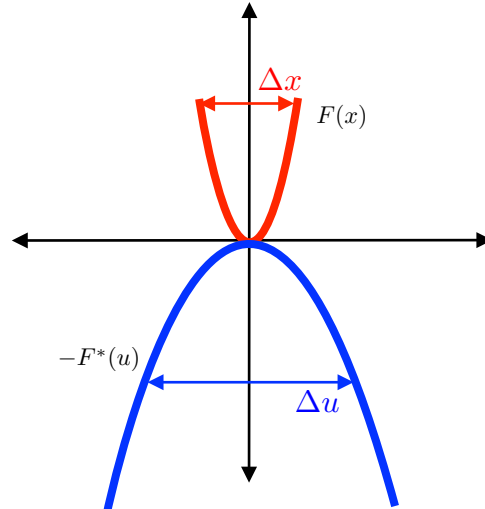
Figure 4.15: Visualization of the convex primal-dual problem where the Legendre transform exhibits a broadening/narrowing of a function similar to that which occurs with a Fourier transform.

4.3 Conclusion

The SPECT reconstruction problem was formulated as an optimization problem, where the objective is to obtain solutions that minimize the D_{KL} data divergence according to equation 4.14. By introducing a penalty term $P(f)$ to the objective function, as shown in equation 4.15, the amount of noise back-projected at each iteration is controlled, thereby reducing noise amplification. Furthermore, by choosing a penalty that promotes sparsity in some basis of the solution, the optimization problem may be carried out using far fewer views than required by traditional FBP or statistical methods. If $P(f)$ is convex, existing convex optimization routines may be used to perform the reconstruction/optimization. The Chambolle-Pock method was chosen as the algorithm for performing SPECT reconstructions because it applies broadly to a large class of convex objective functions which can readily incorporate sparse regularizers, and may therefore be suitable to our few-view reconstruction problem. To our knowledge this is

the first use of the CP method for 3D SPECT using a TV penalty. Hence the aim of the next chapter will be to use simulation studies to characterize the use of the CP method with TV-penalized D_{KL} minimization using the 3D multi-pinhole projection model.

Chapter 5

Aim 2: Method Validation and Characterization

5.1 Introduction

This chapter begins by verifying that the matrix methods proposed in chapter 3 are compatible with the Chambolle-Pock algorithm. An idealized problem is considered and metrics used to characterize the convergence behavior of the algorithm are introduced and subsequently used to evaluate the performance of the rotation-based implementation. The methods are first used to reconstruct a simple phantom using L_2 minimization subject to a quadratic penalty. In this case, the Chambolle-Pock method converges to the exact solution if the forward- and back-projector are well matched. The methods are then applied to the same phantom and reconstructed using TV-penalized KL minimization. Finally, the simulations are used to guide the choice of the appropriate control parameters for the CP method.

5.2 Method verification

Before applying the CP method to constrained D_{KL} optimization we must verify that the rotation-based forward- and back-projector operators are matched in the numerical implementation. As indicated in 4.3.2, the back-projector must be the exact conjugate of the forward-projector, otherwise convergence to a solution is not guaranteed. To test the forward- and back-projectors, a simple convex problem is considered in which the data fidelity measure and the smoothness penalty are quadratic. If the back-projection

is the exact conjugate of the forward-projection, and the problem is convex, then the CP

algorithm converges to a solution. Furthermore, if the optimization problem is quadratic,

and the projection data is ideal, the CP method is guaranteed to converge to the exact

solution [12].

5.2.1(a) The SPECT problem

Convex problems are of particular interest because they have unique optimal

solutions [12, 92]. In order to test the convergence of the forward- and back-projectors

two convex problems are solved using the CP method. A convex problem for SPECT is

formulated using the D_{KL} fidelity metric and a convex smoothness constraint. For the

noise penalty, total variation was chosen,

$$\||\nu|\nabla f|\,\|_1 \tag{5.1}$$

where the reconstruction's gradient magnitude is assumed to be approximately piecewise

constant. The convex optimization problem becomes

$$\hat{f} = \arg\min_{f}\{D_{KL}(\mathbf{H}f, g) + \tfrac{\lambda}{\nu}\||\nu\nabla f|\,\|_1\}, \tag{5.2}$$

as described in chapter 4. The CP method applied to the objective in equation 5.2 is then

given by Algorithm 4.2. Additionally, $\sigma = 1/(sL)$ and $\tau = s/L$ were chosen for the step

sizes of the proximal mappings for $s \geq 1$.

5.2.1(b) The quadratic roughness problem

A quadratic problem was also formulated to test the forward- and back-projectors

by making two modifications to the objective function in equation 5.2; 1) the KL

divergence is replaced by the Euclidean distance, and 2) the TV smoothness penalty is replaced with a quadratic penalty. For the quadratic penalty, "quadratic roughness" is used, and is defined as the square of the 2-norm of the gradient magnitude of f, i.e.

$$QR = \|\nabla f\|_2^2 \quad (5.3)$$

Then substituting $\|\mathbf{H}f - g\|_2$ as the measure of data fidelity and equation 5.3 as the smoothness penalty into the equation 5.2 we obtain the quadratic optimization problem

$$\hat{f} = \arg\min_f \{\|\mathbf{H}f - g\|_2 + \tfrac{\lambda}{\nu^2}\| \, |\nu\nabla f| \, \|_2^2\} \quad (5.4)$$

In order to account for these changes, Algorithm 4.2 only needs minimal modifications; line 6 is replaced by

$$y_n' = y_n + \sigma(\mathbf{H}\bar{f}_n - g) \quad (5.5)$$

and line 9 is replaced with

$$\vec{z}_{n+1} = \vec{z}_n' / (1 + \sigma\nu^2/(2\lambda)) \quad (5.6)$$

After repeated trials using a range of values of σ and τ it was determined that the primal step-size $\tau = 1/(sL)$ and dual step-size $\sigma = s/L$ was optimal for this quadratic problem when $s \geq 1$.

5.2.2 Idealized simulation

To test the algorithm a simple 64^3 voxel phantom, similar to the phantom in chapter 4, was created and is shown in Figure 5.1. Since the objective here is to test our implementation of the forward- and back-projectors, the phantom is intended to be simple; sparse in gradient magnitude, with only a few grayscale values, and a vertical

(a)   (b)

Figure 5.1: The simple digital phantom. (a) A top down view of the central slice. (b) The central slice of a side view, as viewed from left to right in (a).

cross-section that minimizes *multiplexing*, or overlap in the projections formed by each pinhole. The region of high activity has a grayscale value of 250, the region of low activity is 0, and the region that encompasses the two has a grayscale value of 50. The imaging geometry simulates an idealized 5 pinhole system depicted in figure 5.2. An 18 cm×18 cm detector array was simulated with 128×128 evenly spaced pixels.
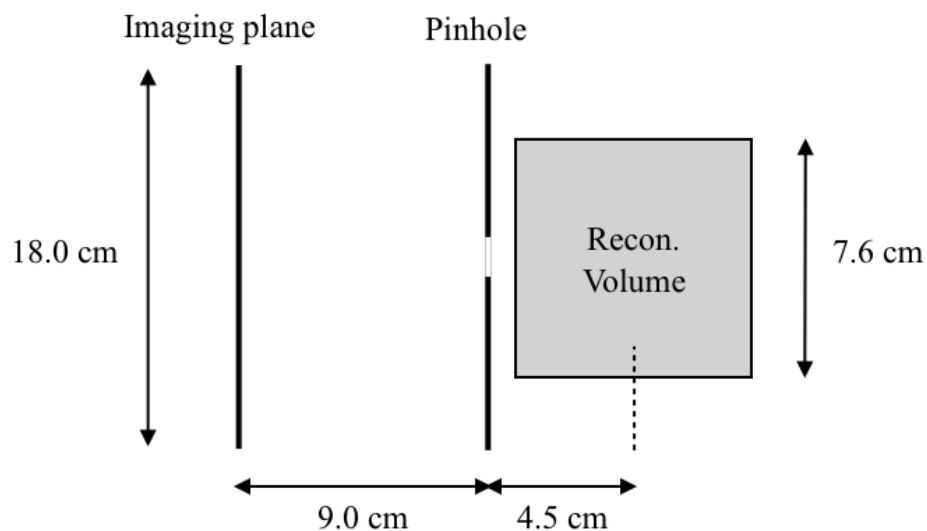
Figure 5.2: The simulated reconstruction geometry used in the simple phantom reconstructions.

5.2.3 Quadratic-penalized L_2-minimization

To begin, noiseless data was simulated using 16 equally spaced projections by applying the forward-projector to the phantom using the idealized 5 pinhole model, shown in figure 2.14(a). Reconstructions were performed using the CP algorithm and varying the penalty term λ between 10^{-4} and 1. The algorithm ran for 2000 iterations with $s = 500$. The central slice of selected reconstructions are shown in figure 5.3. Although the phantoms were recovered using data created by the same forward-projection method used by the algorithm, the pixel-to-pixel distance and the voxel-to-voxel distances were intentionally different to avoid inverse crime [18, 103]. In simulations, inverse crime is "committed" when the inverse problem is posed in such a way that resulting solutions to a problem are better than those expected when applied to a non-simulated problem. For
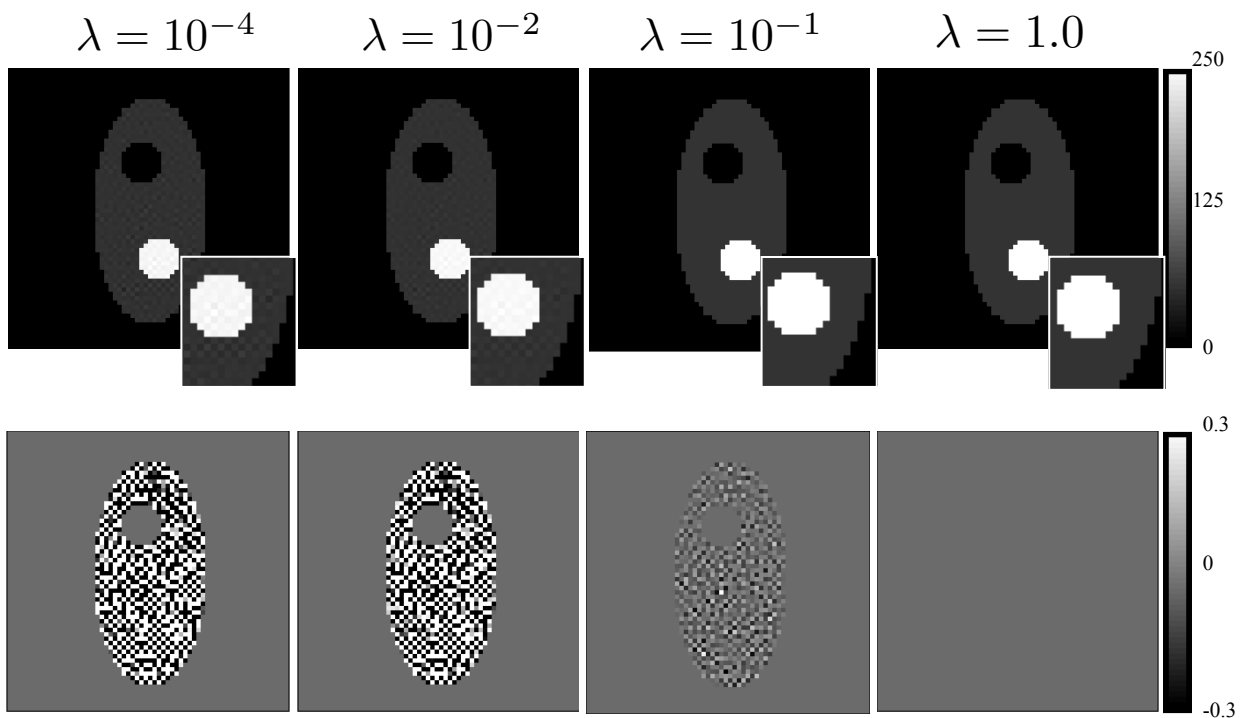
Figure 5.3: Top row: reconstructions of the simple phantom for the quadratic-roughness problem. Bottom row: the difference image between the phantom and the reconstructions.

an $M \times N$ linear system of equations, inverse crime can be committed intentionally by using the same forward operation to solve the problem as was used to simulate the data and by choosing the system dimensions to be $M = N$ [103].

On visual inspection there is little discernible difference between the reconstructions in figures 5.3, and the phantom. The edge features in figures 5.3 are well defined in all cases. The difference image formed by taking the difference between the phantom and the reconstruction are shown below each reconstruction in figure 5.3. The differences between the phantom and reconstructions were less than 0.12% of the maximum grayscale of the phantom (250) using $\lambda = 10^{-4}$ and down to 10^{-4} % for $\lambda = 1$.
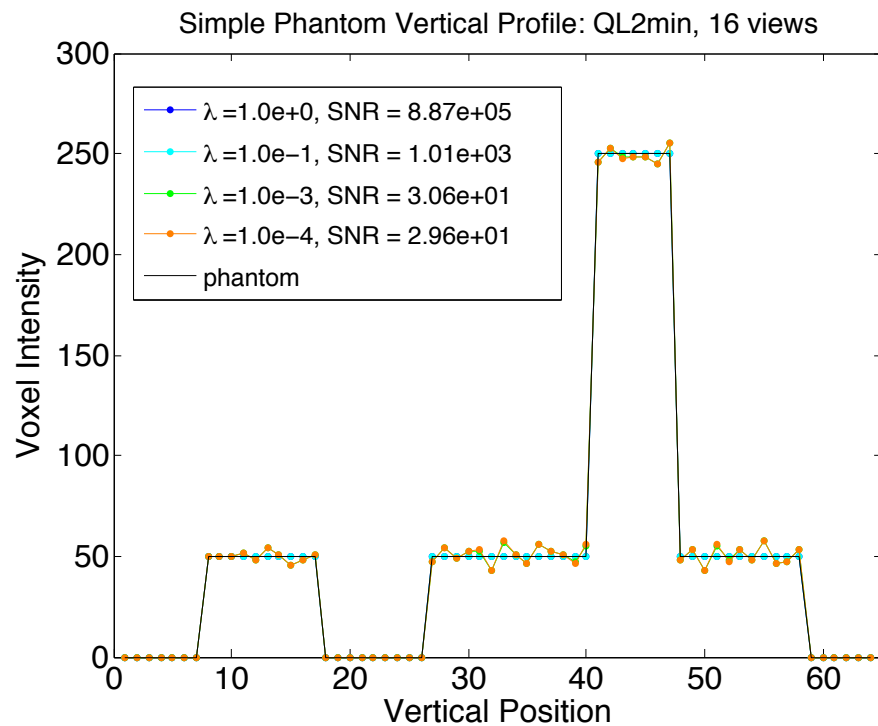
Figure 5.4: The vertical intensity profiles of the reconstructions compared to the phantom.

A plot of the voxel intensities along a vertical profile through the central slice is shown in figure 5.4 for the reconstructions above. In this case, with no noise, we define an objective fidelity measure as a signal-to-noise ratio (SNR)

$$SNR = \sqrt{\frac{\sum\limits_{i=0}^{M-1} \hat{f}_i{}^2}{\sum\limits_{i=0}^{M-1} \left(\hat{f}_i - f_i\right)^2}} \quad . \tag{5.7}$$

where f is the true phantom $\hat{f}$ is the reconstruction. Figure 5.4 reports the SNR and shows that data fidelity increases with λ, with excellent reconstructions for $\lambda = 0.1$ and $\lambda = 1$.

In order to further investigate the performance of the algorithm, we investigated the root mean square error (RMSE) at each iteration. Results plotted in figures 5.5(a) and 5.5(b), show the data RMSE (L_2 norm) between the simulated data and the pseudo-data and the object RMSE for each λ value at each iteration. The data error decreases monotonically for all λ, verifying that our implementation does minimize the data error. Likewise, the object RMSE also decreases monotonically. As expected, based on the results of the reconstructions in figure 5.3, the RMSE between the solution and the phantoms for $\lambda = 10^{-1}$ and $\lambda = 1$ indicates good agreement after 2000 iterations.

Convergence was also measured by examining the behavior of the partial primal-dual gap at each iteration as well as the relative size of the update calculated for each intermediate solution. As mentioned in chapter 4, the partial primal dual gap (*PPDG*) is the difference between the primal and dual objective functions. In the ideal case, *PPDG* is zero at the optimum solution, so we would expect to approach zero if our implementation is converging to the optimal solution. Furthermore, Chambolle and Pock
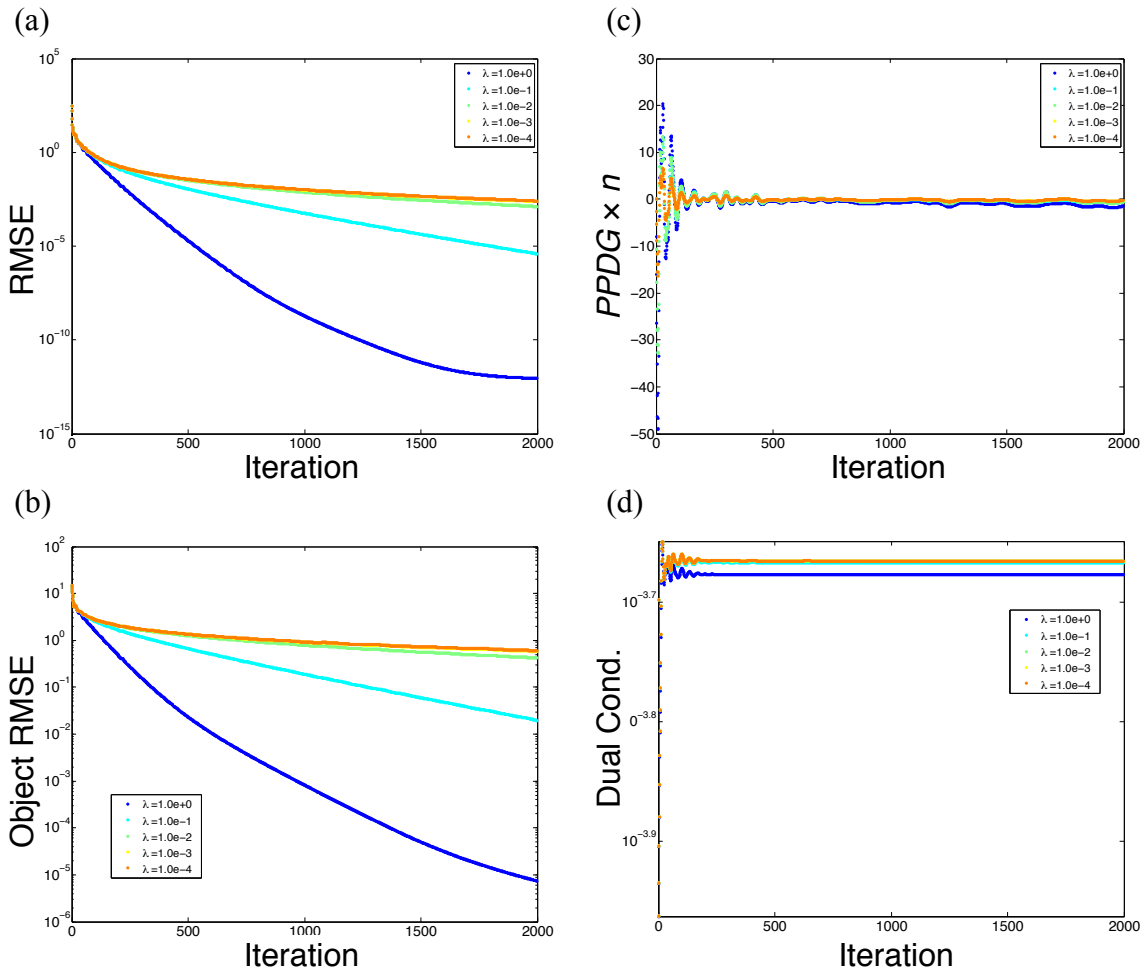
Figure 5.5: The convergence metrics evaluated at each iteration; (a) data RMSE, (b) object RMSE, (c) *PPDG × n*, and (d) dual-condition for 16 view angles.

demonstrate their method will converge to the saddle-point solution, at a rate of $O(1/n)$ [17], where *n* is the number of iterations. To demonstrate that the *PPDG* is approaching zero faster than $1/n$ the *PPDG × n* is plotted at each iteration for the entire range of smoothing parameters in figure 5.5(c). For all values of λ, *PPDG × n* falls between -1.57 and 0.013 after 500 iterations.

An additional convergence metric comes from the update step in line 10 from the CP algorithm 4.2, $f_{n+1} = f_n - \tau(\mathbf{H}^T y_{n+1} + \nu \nabla^T \vec{z}_{n+1})$, where the size of the update to f, known as the *dual-condition*, is given by

$$\text{dual cond.} = \|\mathbf{H}^T y_{n+1} + \nu \nabla^T \vec{z}_{n+1}\|_2 \tag{5.8}$$

The 2-norm was used here since it is a fast metric to calculate on the fly using the GPU. If the algorithm converges, the dual-condition must approach zero. Although the dual-condition using the 2-norm measure is not monotonically decreasing for the quadratic problem it does decrease in the case of TV-penalized KL-minimization as shown in section 5.2.3(b). Figure 5.5(d) shows rapid settling of the dual-condition, i.e. the dual-condition was less than $10^{-3.6}$ after 500 iterations.
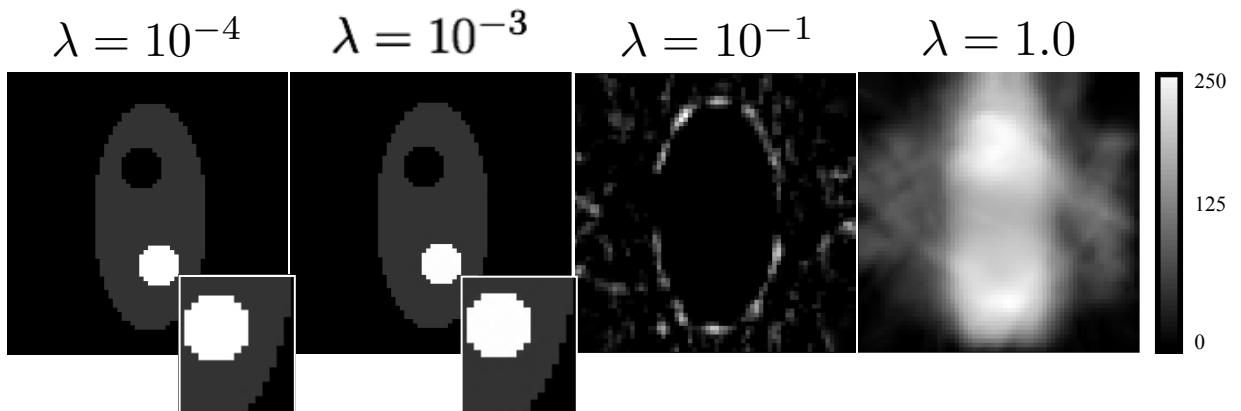
Figure 5.6: Reconstructions of the simple phantom using TV penalized KL minimization. The inset magnifies the hot spot in the case of $\lambda = 10^{-4}$ and 10^{-3}.

5.2.4 TV-penalized KL-minimization

The ideal phantom simulation was repeated using TV-KL minimization (algorithm 4.2) with results shown in figure 5.6. As in the case of the quadratic roughness problem, reasonable reconstructions were obtained using $\lambda = 10^{-4}$ and $\lambda = 10^{-3}$. However, when λ was increased to 10^{-1} and 1 the reconstructions were quite degraded, unlike the quadratic problem above.

Figure 5.7 shows a plot of the vertical voxel intensities and the corresponding SNR of the reconstructed central slice, indicating that for $\lambda = 10^{-4}$ and 10^{-3} the CP algorithm performs well using TV-KL minimization.
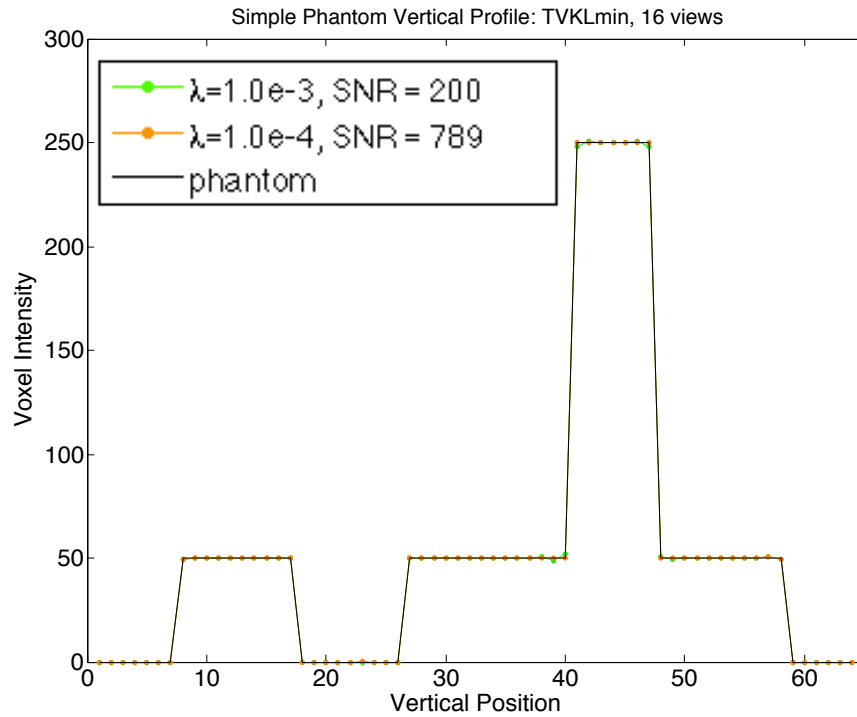
Figure 5.7: The vertical intensity profiles of the TV-KL reconstructions compared to the phantom.

Performance metrics described above are shown in figure 5.8. For the CP method using TV-KL smoothness, Kullback-Leibler data error (D_{KL}) and the object RMSE decrease monotonically only for $\lambda = 10^{-4}$ and 10^{-3}. Although the dual-condition decreased monotonically for $\lambda = 10^{-4}$ and $\lambda = 10^{-3}$, in general the partial primal-dual gap did not approach zero faster than $1/n$, unlike the quadratic problem, suggesting that the method was converging but at a slower rate.
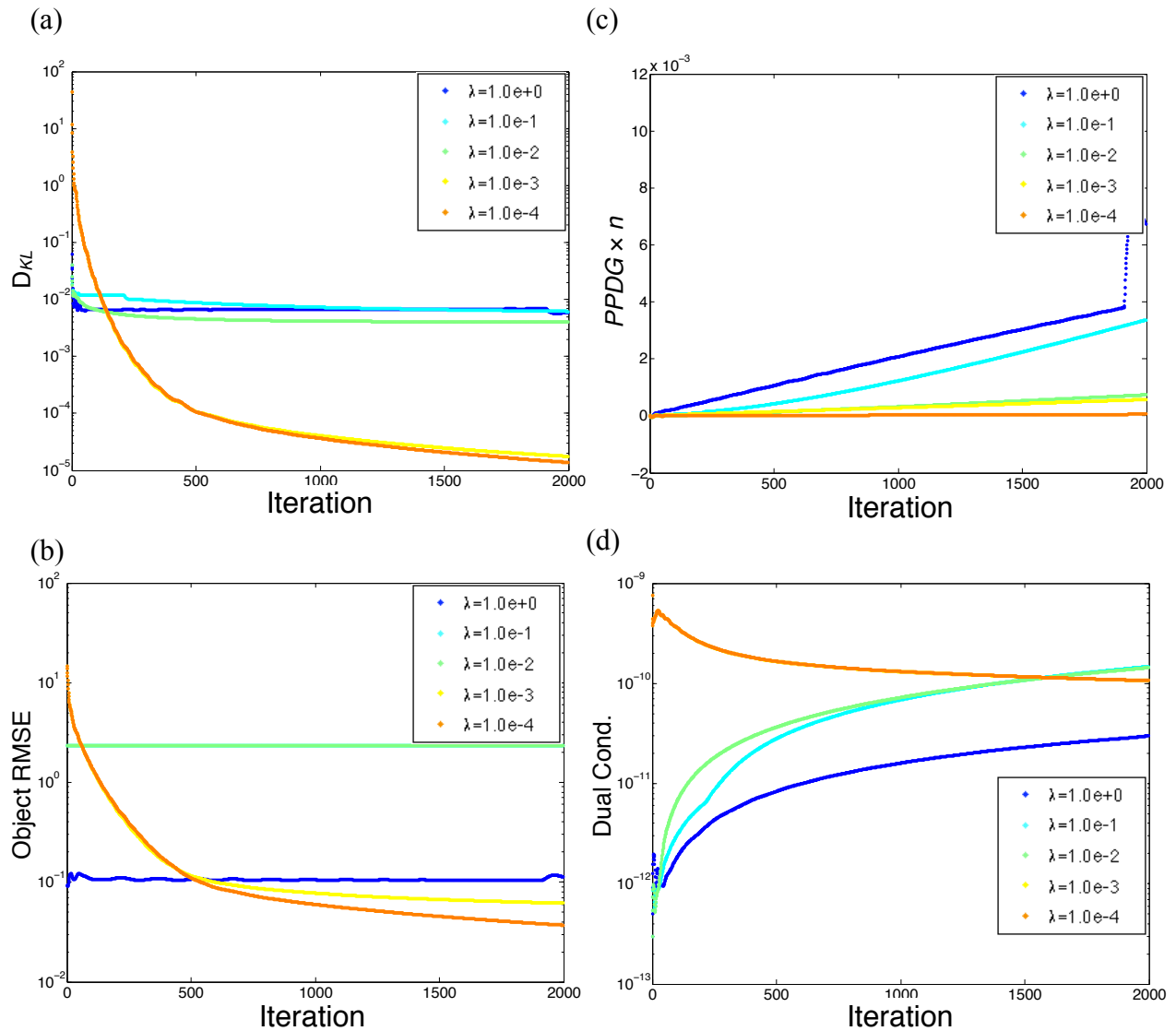
Figure 5.8: TV-KL minimization: the convergence metrics evaluated at each iteration; (a) data error, (b) object RMSE, (c) $PPDG \times n$, and (d) dual-condition for 16 view angles.

5.2.5 Reconstruction of a smooth phantom

TV penalized minimization problems seek solutions that are piecewise constant, however this will not be the case for laboratory specimens in general. Thus, we examined the ability of the CP method to yield accurate reconstructions even when applied to

smoothly varying objects. We reconstructed an object whose gradient magnitude is not sparse and computed the resulting convergence metrics using TV-KL minimization.

The digital phantom of figure 5.1 was modified by applying a Gaussian filter with mean 0 and standard deviation of 1 pixel to obtain the smoothed phantom whose central slices are shown in figure 5.9. Then the forward projector was applied to the phantom to generate data as discussed in section 5.2.3. In the context of CT, the ratio of nonzero elements in the sparse basis to the number of view angles is used as a relative measure of "sparseness" [52]. This ratio was 289 for the simple phantom compared to 2272 for the smooth phantom using TV.
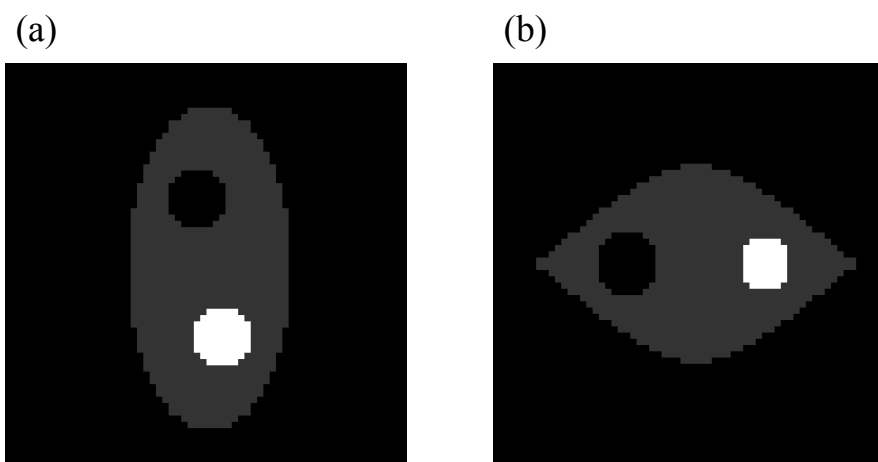
(a)   (b)
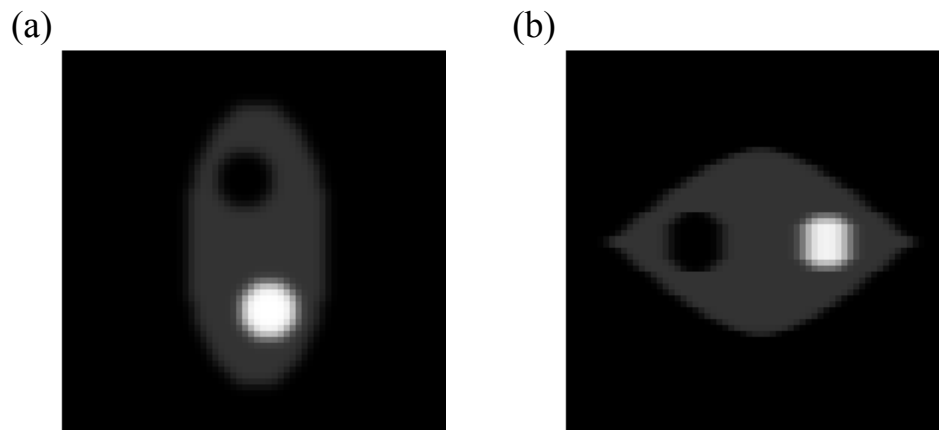
Figure 5.9: The smoothed digital phantom. (a) A top down view of the central slice. (b) The central slice of a side view, as viewed from left to right in (a).

The reconstructions obtained from the smoothed phantom are shown in figure 5.10, and demonstrate good agreement with the actual phantom for $\lambda = 10^{-4}$ and 10^{-3}. This is confirmed in the cross-sectional profile shown in figure 5.11. Convergence metrics for the reconstructions are shown in figure 5.12. Each metric for the smooth phantom follows the same general trend as that obtained using the simple phantom. For $\lambda = 10^{-4}$ and $\lambda = 10^{-3}$ the D_{KL} and the object RMSE are monotonically decreasing with each
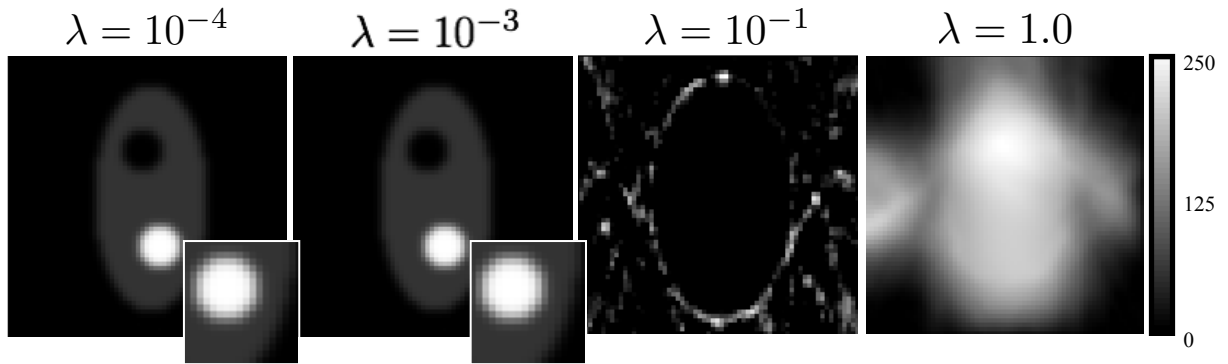
Figure 5.10: Reconstructions of the smoothed phantom using TV-KL minimization.
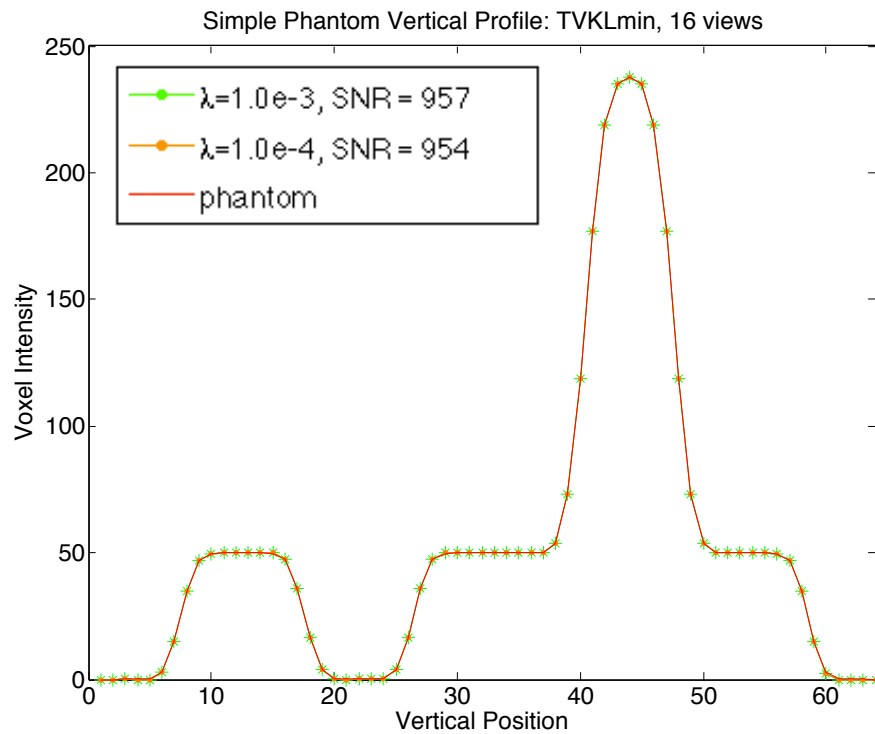
Figure 5.11: Profiles of the smoothed phantom using TV-KL minimization.

iteration, and the *PPDG* demonstrates empirical convergence over 4000 iterations. The dual-condition, again, is monotonically decreasing only when $\lambda = 10^{-4}$ and 10^{-3}. Overall, the CP algorithm using TV penalized KL minimization was successful in reconstructing this smooth phantom in the ideal case.
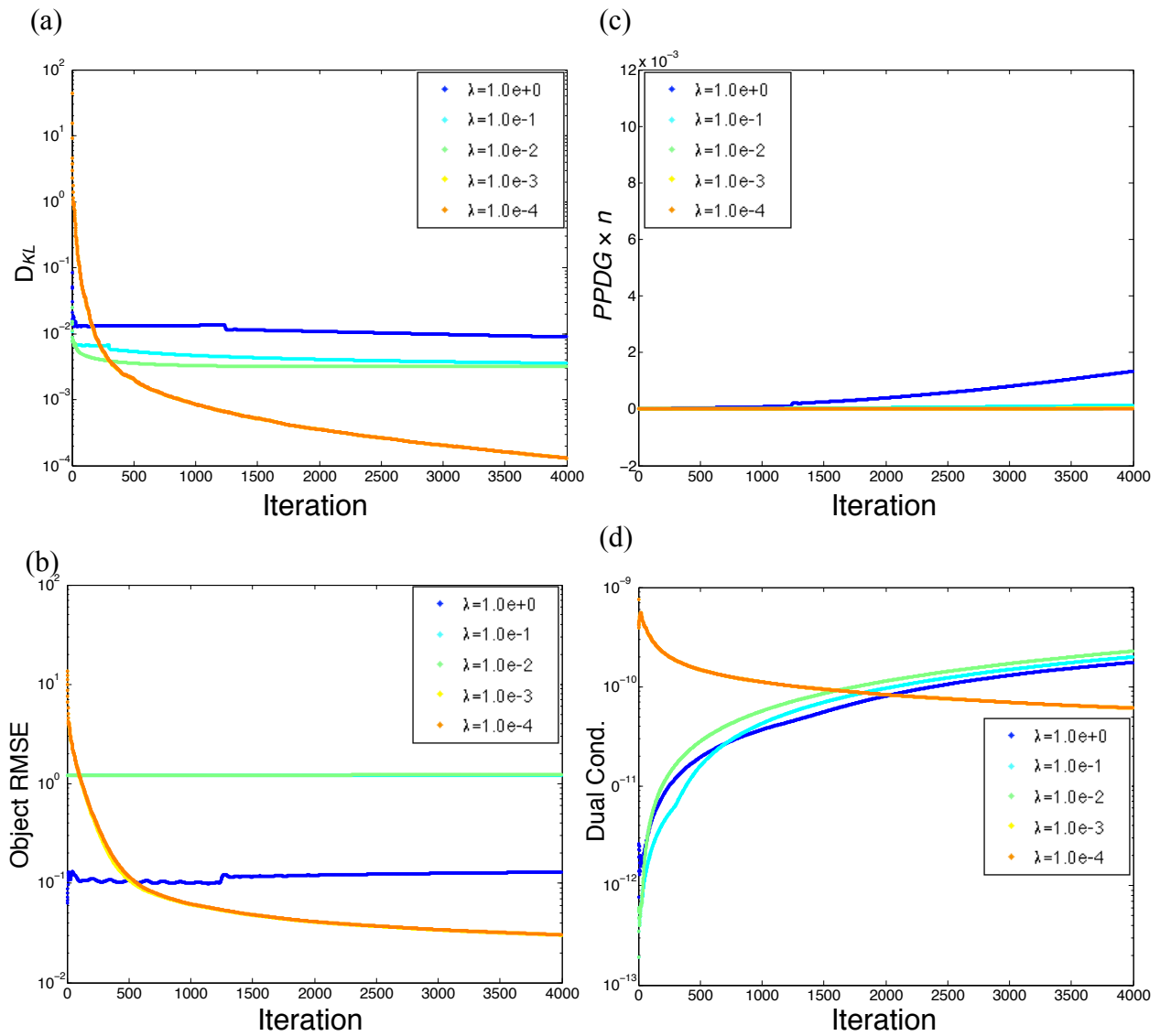
Figure 5.12: Smooth phantom: the convergence metrics evaluated at each iteration; (a) data error, (b) object RMSE, (c) *PPDG × n*, and (d) dual-condition for 16 view angles.

5.3 Simulations using a digital lung phantom with noise

Unlike the ideal reconstruction problem, experimental data is blurred due to the finite size of the pinhole and contains noise due to the emission and detection process. The task is then to investigate CP performance under more realistic conditions and to determine appropriate design parameters for the CP algorithm for data obtained in our

laboratory. A digital rat-lung phantom was created that simulates expected activity distributions, the imaging geometry was adjusted to be consistent with our imaging system, Poisson noise was incorporated into the projection data, and reconstructions with a variable number of view angles were considered. We examined a range of smoothing parameter values λ and scaling parameter values s that included those found to be optimal for the experimental data in chapter 6.

5.3.1 The lung phantom

A digital 3D phantom that simulates the rat lung anatomy and activity distribution of our imaging problem was created. A typical micro-CT rat image was obtained in our laboratory to provide anatomical boundaries and details [82]. Distributed heterogenous activity in the lung regions accompanied by background activity in the surrounding tissue was simulated through a series of digital image processing steps, as shown in figure 5.13.

In figure 5.13, the lung region was isolated by segmenting out the air, or low density regions, of the CT reconstruction [82], resulting in a binary mask containing the lungs alone (figure 5.13(b)). A heterogenous activity distribution was simulated within the lung tissue by adding random Gaussian noise with mean 0 and variance 0.5 to the lung mask, (figure 5.13(c)) then blurring the lung region by applying a gaussian filter with mean 0 and standard deviation of 0.5 pixels (figure 5.13(d)). Activity in the rest of the body outside of the lungs was simulated by first segmenting the region surrounding the lungs, then setting the activity level to 3% of the lung region, adding Gaussian noise and blurring, as before, (figure 5.13(e)). Finally, the large blood vessels were reinserted into the lung regions and correspond to the black spots in figure 5.13(f). The final result
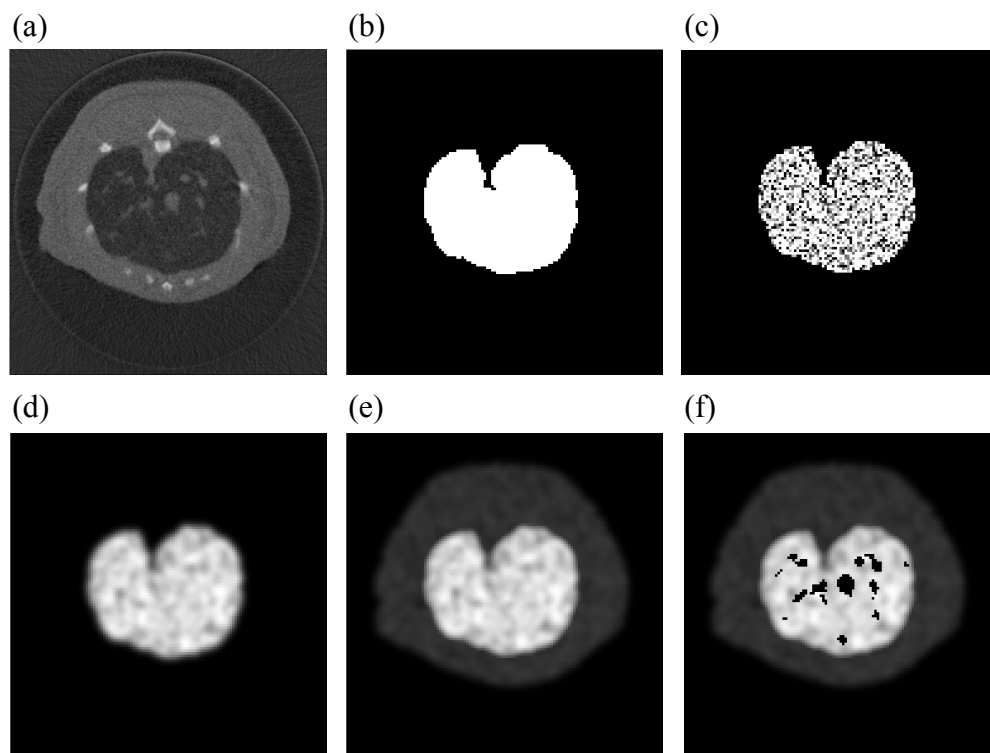
Figure 5.13: Creating the 3D digital lung phantom; (a) Original axial CT reconstruction, (b) binary lung mask, (c) Gaussian noise applied to the lung region, (d) blurring to simulate heterogenous uptake, (e) inclusion of the surrounding tissue, (f) reinsertion of the large blood vessels.

is shown in figure 5.14 at various slice levels through the phantom. Additionally,

phantoms with mean voxel intensities within the lungs ranging from 160 to 1280 were

generated and are shown in figure 5.15.
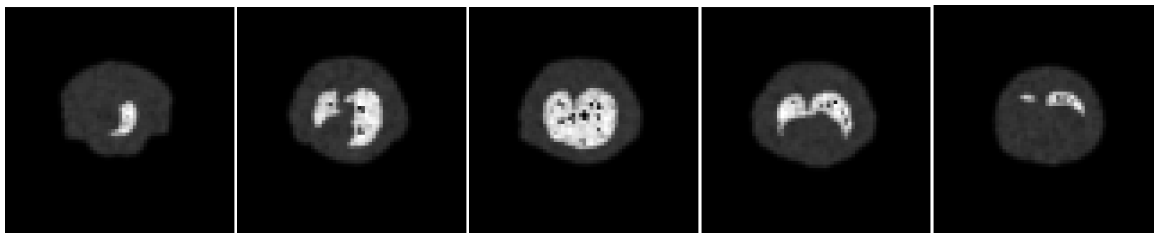
Figure 5.14: Axial slices through the rat lung phantom. The images pan from the top of the lung, at the left, to the bottom of the lung, on the right.
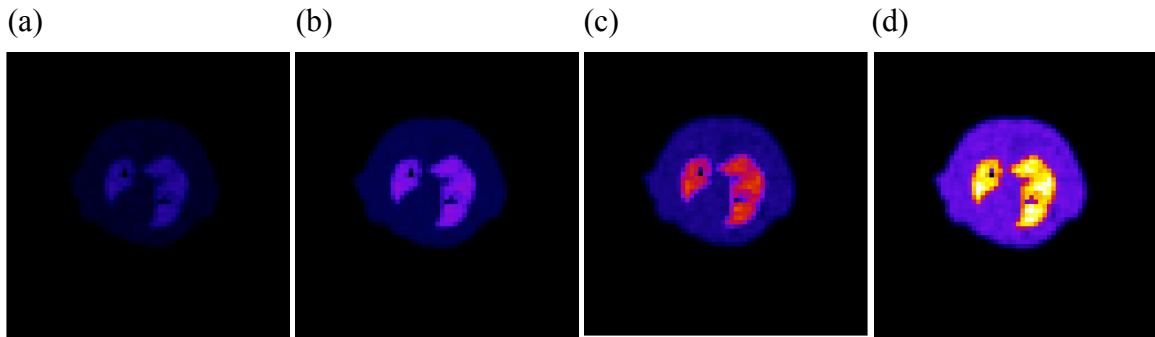
Figure 5.15: The lung phantom simulating different mean activity levels in the lungs: mean intensities of (a) 160 , (b) 320, (c) 640, and (d) 1280 in the lung region.

5.3.2 The effect of changing the smoothing parameter λ

We investigated the effect of the magnitude of the smoothing parameter λ in our CP reconstruction method. We used an optimal scaling parameter, $s = 100$, since it allows us to better observe the effect of λ on the reconstructions. For each value of λ, reconstructions were performed by simulating 3, 6, 15, and 60 view angles with a collimator consisting of 5 pinholes, each with a 2 mm diameter, consistent with our laboratory's imaging geometry (see chapter 6). The effective "acquisition time" was kept constant for the 3, 6, 15, and 60 view data sets. To further simulate data consistent with experimental data, 5% Poisson noise [8] was applied to the phantom projections. Since Poisson noise is dependent on the amplitude of the signal, adjusting the variance in the projections changes the amplitude of the projections as well. To preserve the mean amplitude of the projection and adjust the variance to a prescribed level, Poisson noise was applied according to [58], and is summarized in algorithm 5.1. For each pixel in the projection g_0, a poisson random number, $g_1[i]$, is generated using mean $g_0[i]$ (line 3). A new mean is generated by dividing the variance of $g_0[i]$ by the desired amount of

fractional noise to be applied to the projection (line 4). The standard deviation is adjusted

to the new noise level (line 5), and finally, the mean of the random variable $g_1[i]$ is

adjusted back to the original mean, $g_0[i]$ (line 6).

$$
\begin{aligned}
&1: \quad \text{INPUT} : \beta = 1.0 - 0.05 \\
&2: \quad \textbf{for } i < M \\
&3: \qquad g_1[i] = \text{poissRnd}\,(g_0[i]) \\
&4: \qquad \text{newMean} = \sqrt{g_0[i]}/\beta \\
&5: \qquad g_1[i] + = \text{newMean} - g_0[i] \\
&6: \qquad g_1[i] = g_0[i] \cdot g_1[i]/\text{newMean} \\
&7: \quad \textbf{end}
\end{aligned}
$$

Algorithm 5.1: Applying 5% Poisson noise to a projection g_0.

Phantom reconstructions were performed using the CP algorithm applied to the

TV-KL minimization problem using 2000 iterations and λ ranging from 0.01 to 10.

Figure 5.16 shows the central slice of the CP reconstructions, as well as the

corresponding reconstruction obtained using 30 iterations of MLEM. Unless otherwise

noted, the following results were obtained from the phantom with a mean lung intensity

of 640. As λ increased, the reconstructions appeared blurrier. As the number of view

angles increased there was a reduction in artifacts outside the body of the "rat" and the

features of the phantom were better resolved. For small λ, more activity was assigned to

regions outside the boundaries of the phantom since the data was undersampled. For

MLEM, the visual image quality improved with the number of view angles, as expected.

Figure 5.16: Reconstructions: each row represents 3, 6, 15, and 60 view angles. The first four columns are the CP reconstructions for a given smoothing parameter and $s = 100$. MLEM is shown in the last column for comparison. 5% noise.

To quantitatively assess the reconstructions the correlation coefficient (CC) between the reconstruction and the phantom was calculated according to

$$CC = \frac{\sum\limits_{i=0}^{M-1} \left(\hat{f}_i - \bar{\hat{f}}\right)\left(f_i - \bar{f}\right)}{\sqrt{\sum\limits_{i=0}^{M-1} \left(\hat{f}_i - \bar{\hat{f}}\right)^2 \sum\limits_{i=0}^{M-1} \left(f_i - \bar{f}\right)^2}} \tag{5.9}$$

where $\hat{f}$ is the reconstruction, f is the phantom, $\bar{\hat{f}}$ is the mean of the reconstruction, and

$\bar{f}$ is the mean of the phantom. The CC was used as a measure of reconstruction fidelity

[44]. SNR and CC were calculated from the full reconstruction volume and are reported

in Table 5.1. Overall, CC was only slightly larger for CP with $\lambda = 1$ compared to MLEM.

In this example, SNR is higher for the CP method than for MLEM for all view angles.

Additionally, improved SNR was observed for reconstructions with $\lambda = 1$ over $\lambda = 0.01$

for 3 and 15 views.

3 views	SNR	CC
$\lambda = 0.01$	2.37	0.915
$\lambda = 1$	2.75	0.945
MLEM	1.50	0.942
15 views	SNR	CC
$\lambda = 0.01$	3.73	0.962
$\lambda = 1$	3.92	0.970
MLEM	3.28	0.960
60 views	SNR	CC
$\lambda = 0.01$	4.67	0.980
$\lambda = 1$	4.06	0.971
MLEM	3.62	0.962

Table 5.1: SNR and CC for the full reconstruction volumes.

Figure 5.17 shows that for a fixed number of views, the mean of the central slice

profile obtained from the CP method did not appear to be sensitive to λ, however, more

variation was observed in the intensity profiles with $\lambda = 0.01$. Careful examination of the

reconstructed images in figure 5.16 reveals that more activity was assigned to regions

outside the phantom boundaries using $\lambda = 1$, leading to lower estimates of the peaks
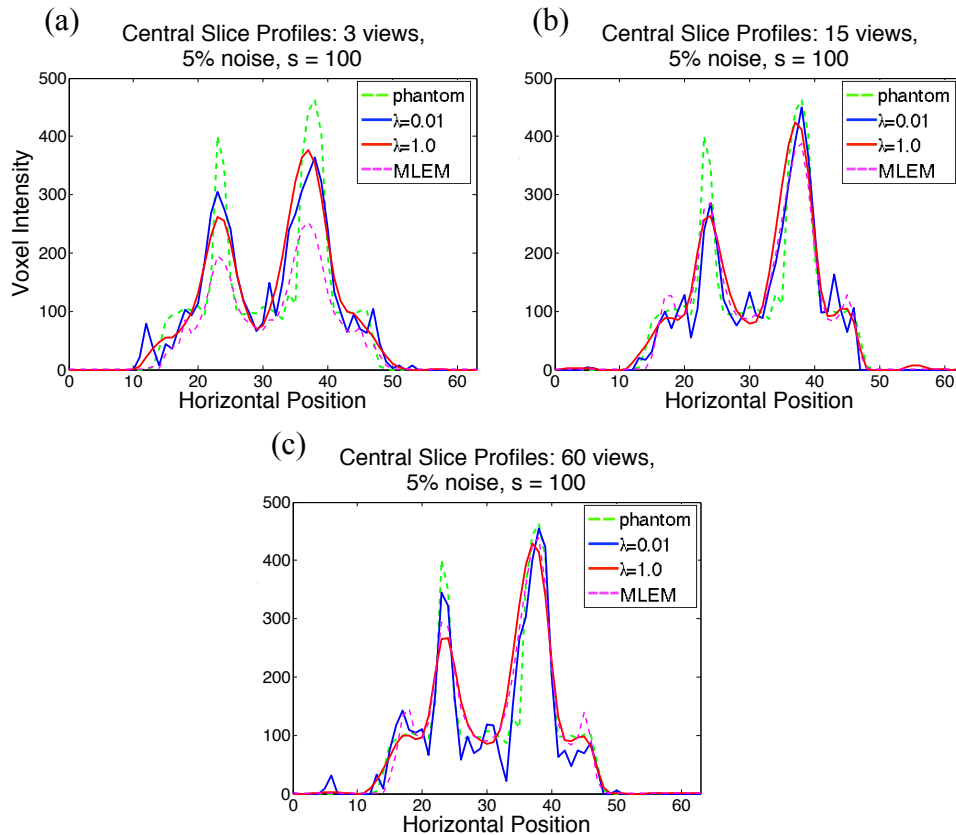
Figure 5.17: Horizontal intensity profiles taken from the central slice: (a) 3 views, (b) 15 views, and (c) 60 views. For $s = 100$ and 5% noise. The true profile is shown in green, and MLEM is shown in magenta.

intensities compared to $\lambda = 0.01$. Figure 5.17(a)-(c) also shows that increasing the number of view angles, from 3 to 60, provided better estimates of the peak intensities for both the CP method and MLEM. When only 3 view angles were used, the peak intensities obtained using the CP method were closer to the true peak intensities than those obtained using MLEM.

To assess the convergence behavior of the CP method applied to this lung phantom, convergence metrics were computed and are shown in figure 5.18. The trends for each metric appear independent of the number of view angles used in the reconstruction. The data error and the dual-condition decreased monotonically, and as the
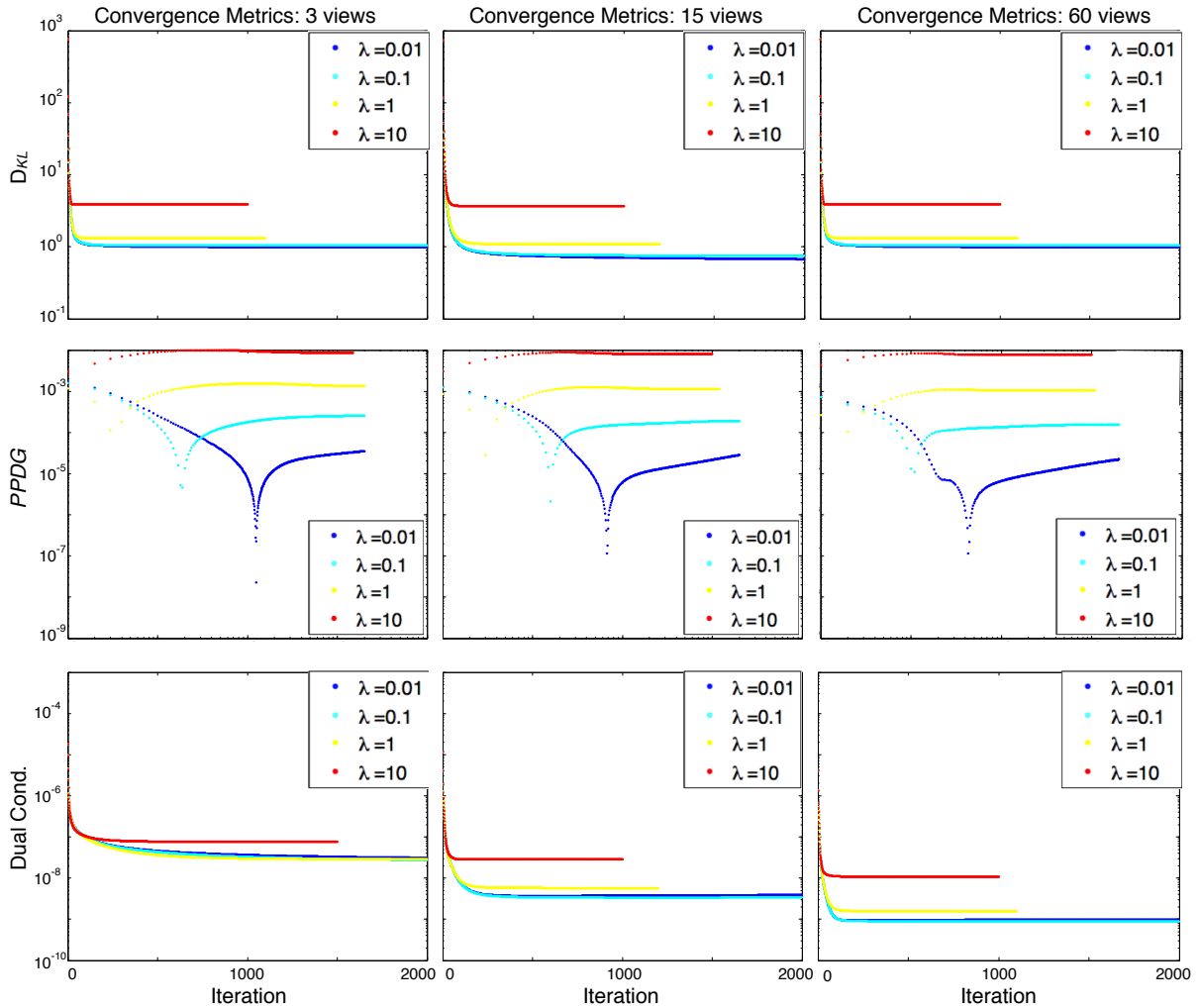
Figure 5.18: The CP convergence metrics: The rows correspond to the data D_{KL}, the *PPDG*, and the dual-condition. The columns display the results for 3, 15, and 60 view angles. In all cases, $s = 100$ with 5% noise applied to the data.

number of views increased the data error and the dual-condition decreased at a faster rate. *PPDG* also decreased with iteration, in general, but at a much slower rate than in the ideal case. In figure 5.18 the partial primal-dual gap was plotted on a log-log scale so that the spike features indicate where *PPDG* crosses 0. After the zero crossing, the rate at which *PPDG* decreased dropped off considerably.

For all view angles the smaller the smoothing parameter the smaller the data error after 2000 iterations, as expected. Likewise, the *PPDG* and the dual-condition were

smaller for $\lambda < 10$ after 2000 iterations. However, using smaller λ required more iterations before the *PPDG* reaches the zero-crossing feature. The dual-condition also required more iterations before leveling off for $\lambda < 10$.

Since many SPECT studies are concerned with quantifying the activity within some region of interest (ROI) [50] of the reconstructed volume, we examined the ability of the CP method to accurately recover the mean intensity value within the lung region. Reconstructions were performed on simulated data generated from the phantoms with mean voxel intensities of 160, 320, 640 and 1280. Poisson noise was applied to the simulated data without changing the mean, and 100 data sets were created for each phantom using different noise realizations. Reconstructions were repeated using a range of λ values and number of view angles.

Table 5.2 show the resulting $\overline{\mathrm{SNR}}$ and $\overline{\mathrm{CC}}$ for all 100 realizations of each phantom. Figure 5.19 shows plots of the estimated mean activity from the 100 noise realizations versus the true mean intensity with TV-KL minimization for $\lambda = 1$ with

3 views, $\lambda = 1$			MLEM	
Intensity	$\overline{\mathrm{SNR}}$	$\overline{\mathrm{CC}}$	$\overline{\mathrm{SNR}}$	$\overline{\mathrm{CC}}$
160	2.7	0.94	1.5	0.94
320	2.9	0.95	1.5	0.94
640	2.9	0.95	1.5	0.94
1280	2.9	0.95	1.5	0.94

60 views, $\lambda = 1$			MLEM	
Intensity	$\overline{\mathrm{SNR}}$	$\overline{\mathrm{CC}}$	$\overline{\mathrm{SNR}}$	$\overline{\mathrm{CC}}$
160	4.1	0.97	3.6	0.96
320	4.0	0.97	3.6	0.96
640	4.0	0.97	3.6	0.96
1280	4.1	0.97	3.6	0.96

15 views, $\lambda = 1$			MLEM	
Intensity	$\overline{\mathrm{SNR}}$	$\overline{\mathrm{CC}}$	$\overline{\mathrm{SNR}}$	$\overline{\mathrm{CC}}$
160	4.0	0.97	3.3	0.96
320	3.9	0.97	3.3	0.96
640	4.0	0.97	3.3	0.96
1280	4.0	0.97	3.3	0.96

Table 5.2: Average SNR and CC for the full reconstruction volumes.

varying number of views. Note that in all cases the standard deviation of the mean

activity estimate was less than 0.2% of the mean activity estimate, as indicated by the

small error bars in figure 5.19. This small standard deviation is in part due to the large

lung ROI, and would be expected to increase as the size of the ROI decreases. For each
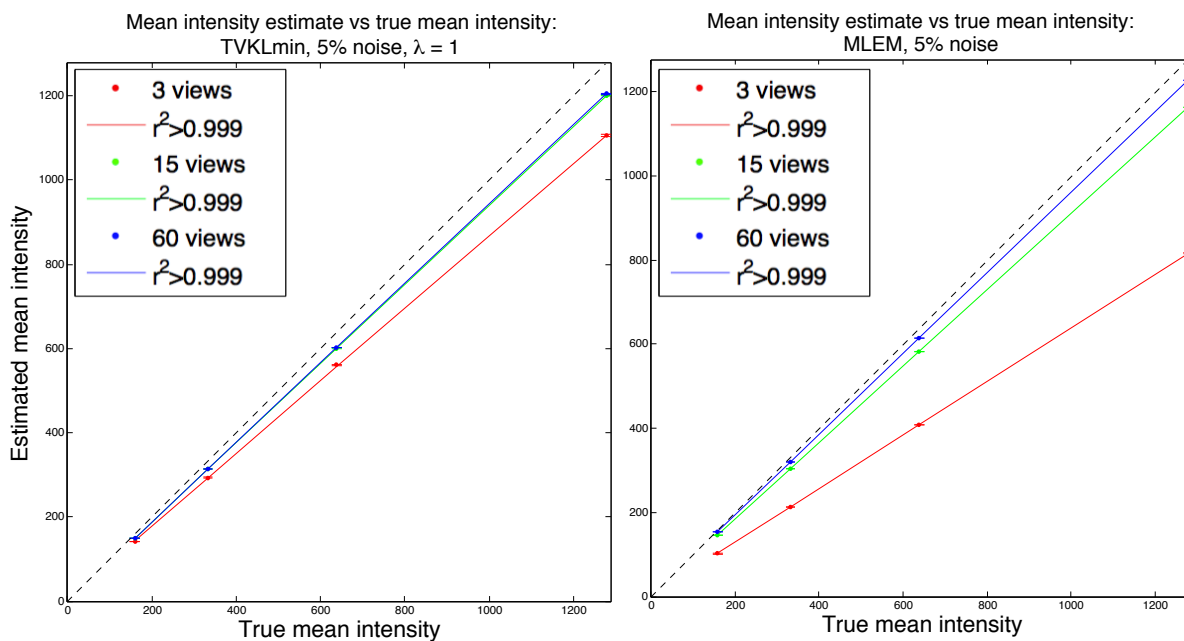
Figure 5.19: Mean intensity estimates obtained from 100 noise realizations plotted against the known intensities in the lung regions of the phantom for $\lambda = 1$. Linear regressions for other λ are summarized in table 5.1.

case a line was fit to the data, and the slopes were compared to 1, where a slope of 1

would be obtained if the reconstruction method perfectly recovered the mean intensity.

Corresponding results for MLEM are shown in the right panel of figure 5.19. The results

are summarized in table 5.3 which shows that the intensity estimates were not sensitive to

the choice of λ and the estimates improved when more view angles were used. The mean

intensity in the lung regions was recovered better with the CP method than with MLEM

when only 3 views were used, for 15 view angles and λ=1, the two methods were roughly

comparable, but for 60 views MLEM outperformed the CP method.

3 views	$\lambda = 0.01$	$\lambda = 0.1$	$\lambda = 1$	$\lambda = 10$	MLEM
m	0.858	0.859	0.862	0.702	0.640
b	16.529	16.346	5.697	0.346	-0.017
r^2	>0.999	>0.999	>0.999	>0.999	>0.999
15 views					
m	0.985	0.980	0.941	0.814	0.911
b	2.442	0.962	-0.198	-0.452	0.005
r^2	>0.999	>0.999	>0.999	>0.999	>0.999
60 views					
m	0.993	0.985	0.943	0.820	0.962
b	0.929	-0.064	-0.353	-0.895	0.008
r^2	>0.999	>0.999	>0.999	>0.999	>0.999

Table 5.3: Linear fits to the mean intensity estimates within the lung regions of the
phantoms presented in figure 5.15.

5.3.3 The effect of changing the scaling parameter *s* on activity estimates

We investigated the effect of changing the scaling parameter using $s = 1, 10, 50,$

and 100 with $\lambda = 1$ and 5% noise. The resulting reconstructions are shown in figure 5.20.

As *s* increased fewer artifacts were observed outside the body region, providing

better resolution of the outermost boundary of the phantom, particularly for 3 view

angles, as shown in the top row. Additionally, as *s* increased the intensity within the lung

boundaries increased, which can be seen by comparing the overall intensities in the
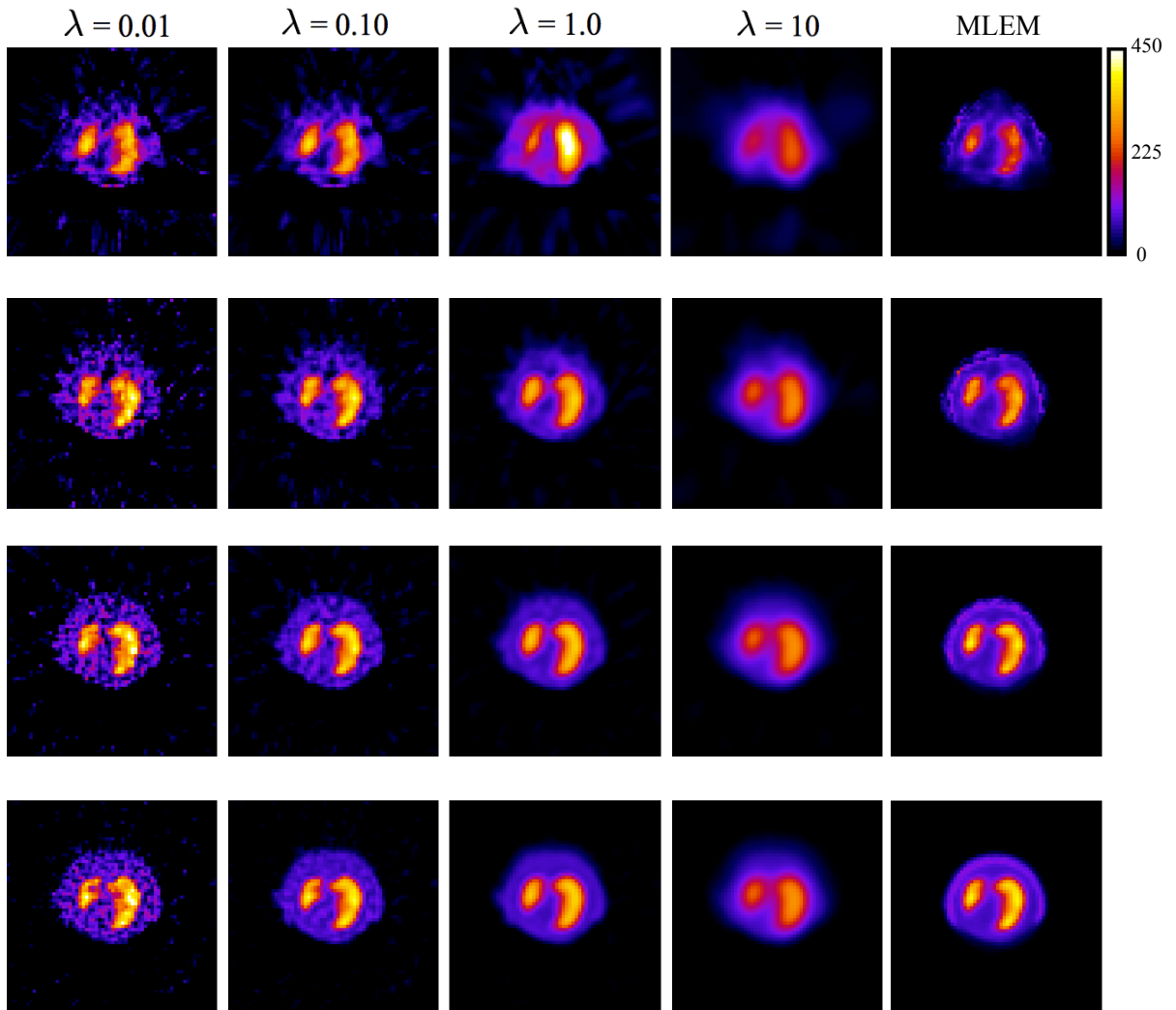
images of row 1 in figure 5.16 to row 1 in figure 5.20.
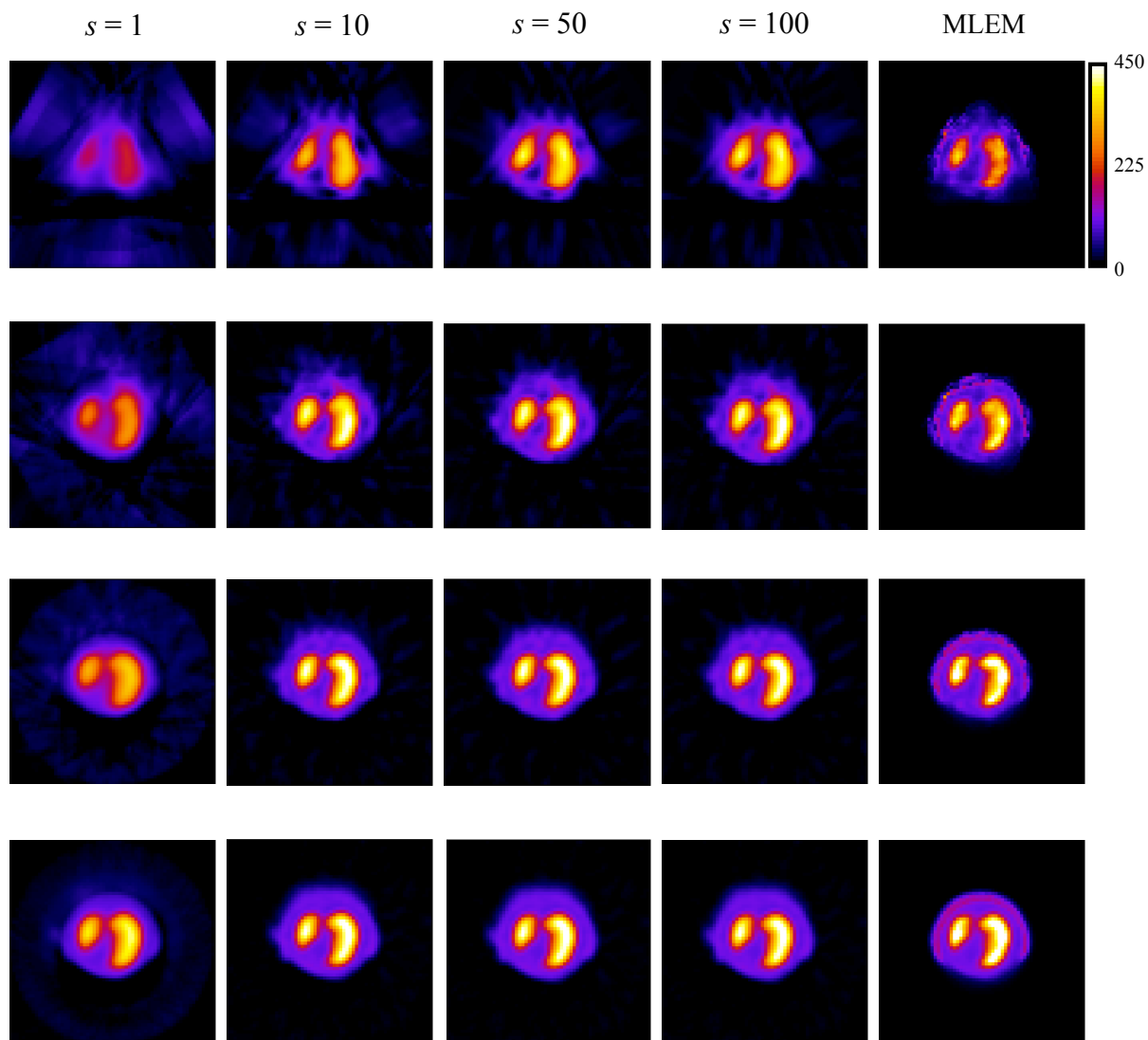
Figure 5.20: Reconstructions: Each row represents 3, 6, 15, and 60 view angles. The first four columns show the CP reconstructions for a given scaling parameter and $\lambda = 1$. MLEM is shown in the last column for comparison. 5% noise.

Figure 5.21 shows central slice profiles obtained using $\lambda = 1$ and $\lambda = 0.01$. The intensity estimates when $s = 1$ and $s = 50$ were quite different, but for $s > 1$, the CP method yielded peak intensities closer to the phantom's than MLEM. For only 3 view angles, MLEM provided a better estimate of the intensity profile than the CP method only in the case $s = 1$. As before, increasing the number of view angles further improves the
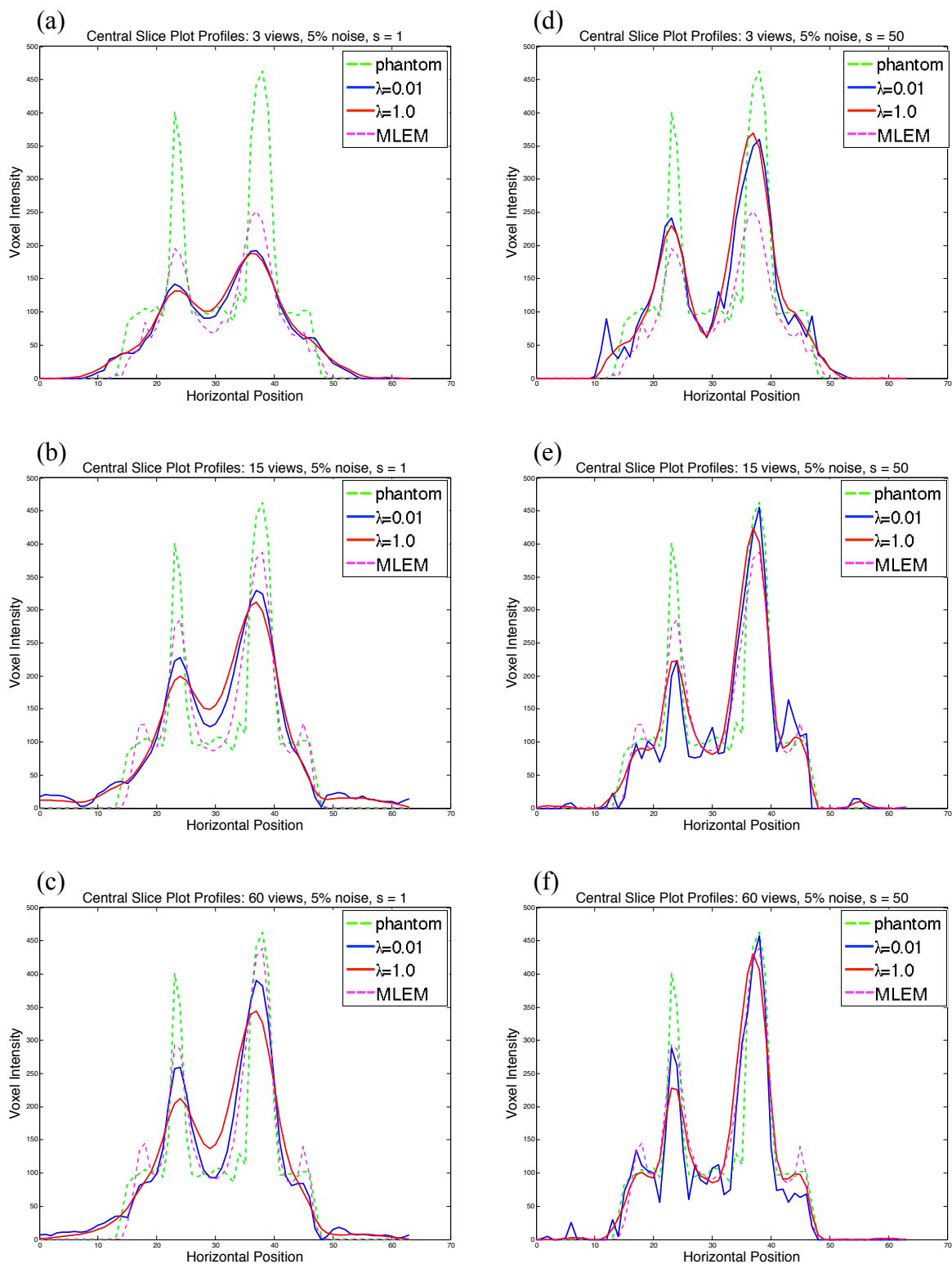
Figure 5.21: Horizontal intensity profiles taken from the central slice for *s* = 1 and *s* = 50. The rows represent 3, 15, and 60 view angles. The true profile is shown in green, and MLEM is shown in magenta.

peak intensity estimates for all values of *s*.

Table 5.4 demonstrates that the SNR and CC are consistent with these observations. The SNR and CC of the reconstructions were improved for all view angles using $s = 50$ compared to $s = 1$.

The effect of *s* on the CP method's ability to estimate the mean intensity within

3 views, $s = 1$	SNR	CC
$\lambda = 0.01$	0.806	0.755
$\lambda = 1$	0.805	0.757
3 vies, $s = 50$	SNR	CC
$\lambda = 0.01$	2.32	0.917
$\lambda = 1$	2.57	0.939
MLEM	1.50	0.942
15 views, $s = 1$	SNR	CC
$\lambda = 0.01$	1.64	0.907
$\lambda = 1$	1.63	0.906
15 vies, $s = 50$	SNR	CC
$\lambda = 0.01$	4.20	0.971
$\lambda = 1$	3.92	0.970
MLEM	3.28	0.959
60 views, $s = 1$	SNR	CC
$\lambda = 0.01$	2.40	0.937
$\lambda = 1$	2.35	0.935
60 vies, $s = 50$	SNR	CC
$\lambda = 0.01$	5.19	0.981
$\lambda = 1$	4.06	0.971
MLEM	3.62	0.962

Table 5.4: Reconstruction SNR and CC for 3, 15, and 60 view angles, using $s = 1$ and 50, and 5 % noise.

the lung boundaries was also investigated. Reconstructions were performed, as described previously, and mean intensity estimates were plotted against the true mean intensity of the phantoms. Figure 5.22, shows the results and the linear fit to the data. Data was

plotted for reconstructions using $s = 1$, 50, and 100 with 3 or 15 views and $\lambda = 1$.
Reconstructions using the CP method with $s = 50$ and $s = 100$, show better intensity estimates over the entire range of simulated activity levels compared to MLEM. The converse is true when using $s = 1$.

The convergence metrics in figure 5.23 show that for 3 view angles the trends are
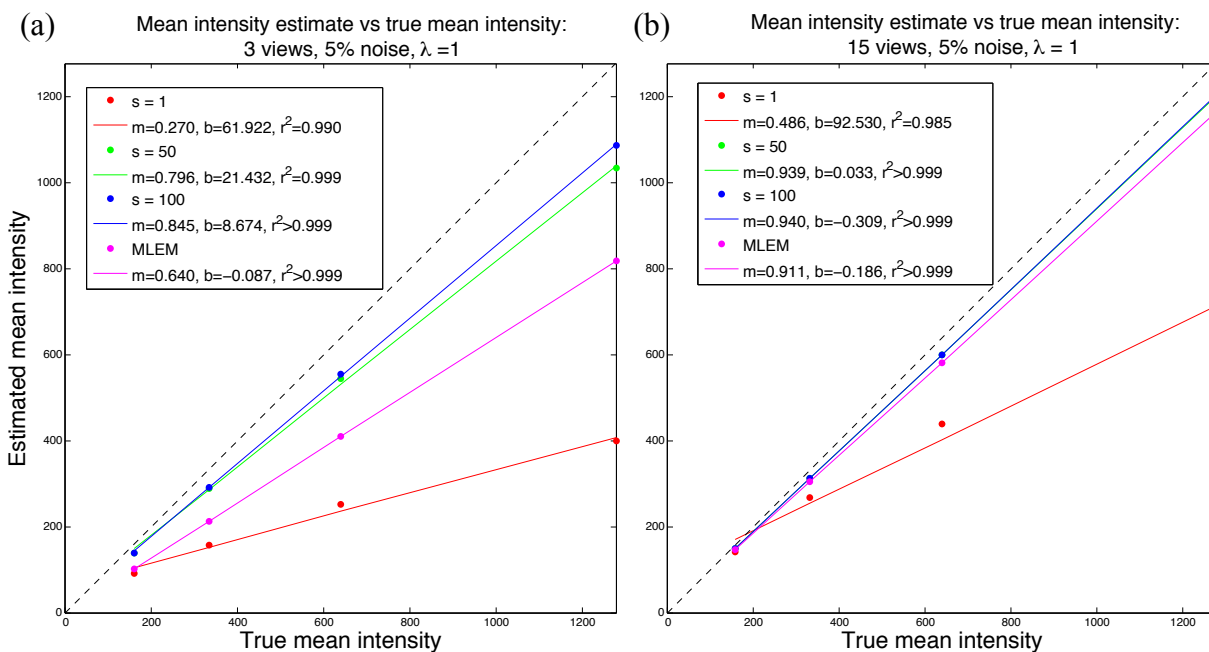
Figure 5.22: Mean intensity estimates in the lung region for $\lambda = 1$: (a) 3 view angles, and (b) 15 view angles.

similar to results obtained by varying λ in figure 5.18. The data error and the dual-condition decreased monotonically for all s and view angles and decayed fastest for $s = 100$. The *PPDG* exhibits the same zero crossing feature as in figure 5.18, and for all λ, the zero crossing was observed with fewer iterations as s was increased. Since s controls the scaling of the primal and dual step-sizes, the results suggest that for 3 view angles, using $\lambda = 1$ and $s = 100$ is more suitable for our reconstruction problem.
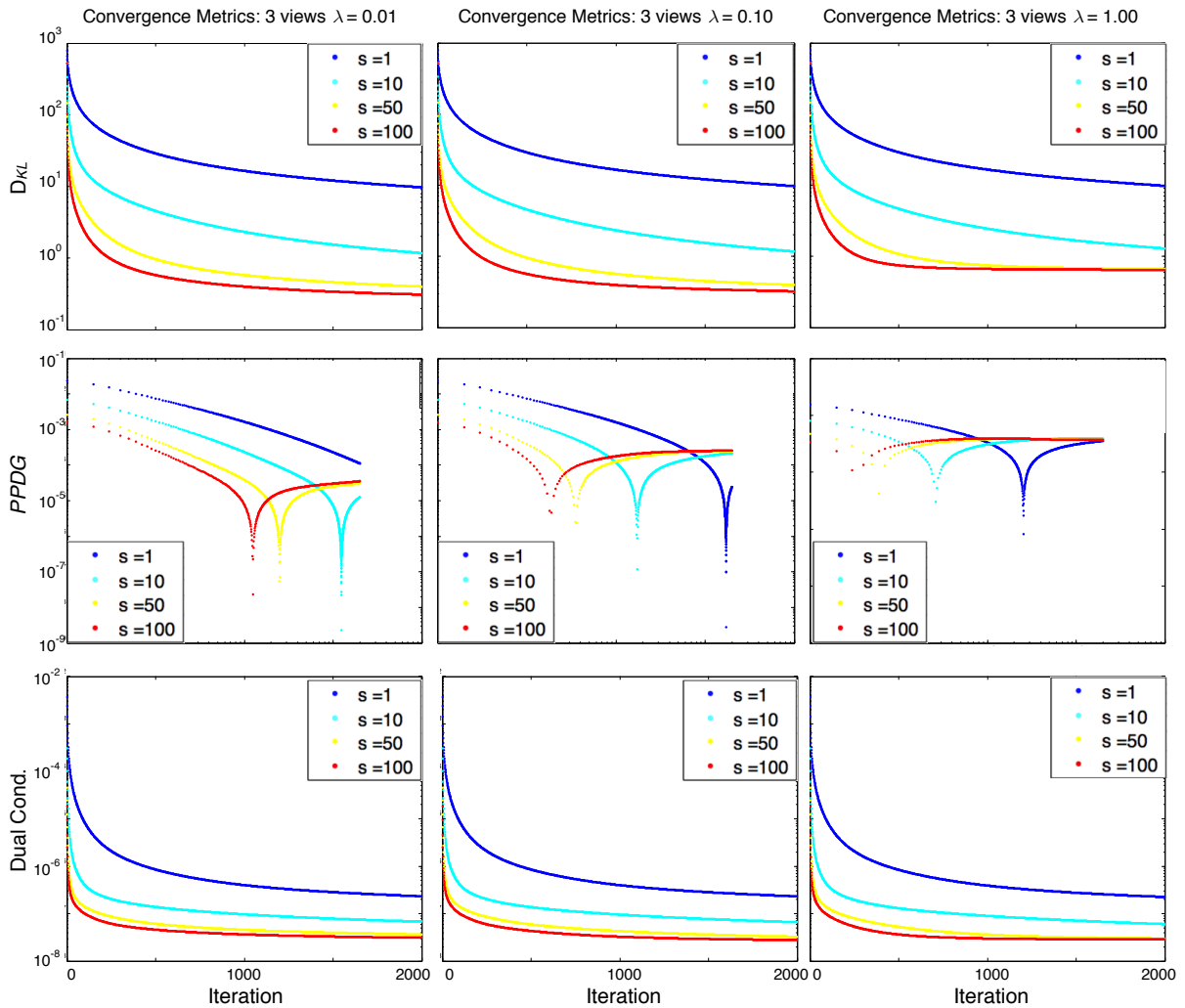
Figure 5.23: Convergence metrics for 3 view angles with $s = 1$, 10, 50, and 100. The columns display the results for $\lambda = 0.01$, 0.10, and 1.00.

5.3.4 TV-penalized reconstructions using the L$_2$ data error constraint (TV-L$_2$)

In this section the effect of changing the data error constraint on the reconstructions was examined. Previously, the optimization procedure used the D$_{KL}$ data fidelity metric because of the Poisson noise model. Here we examined the effect of using L$_2$ data fidelity instead on data simulated with 5% Poisson noise. This was investigated

by reconstructing the 640 lung phantom from section 5.3.1 using $s = 100$, with results shown in figure 5.24.

The resulting reconstructions using TV-L_2 minimization were considerably noisier than those in figure 5.16, even in the case of $\lambda = 10$. However, the reconstructed activity assigned to regions outside of the rat boundary was not distributed over a large number of
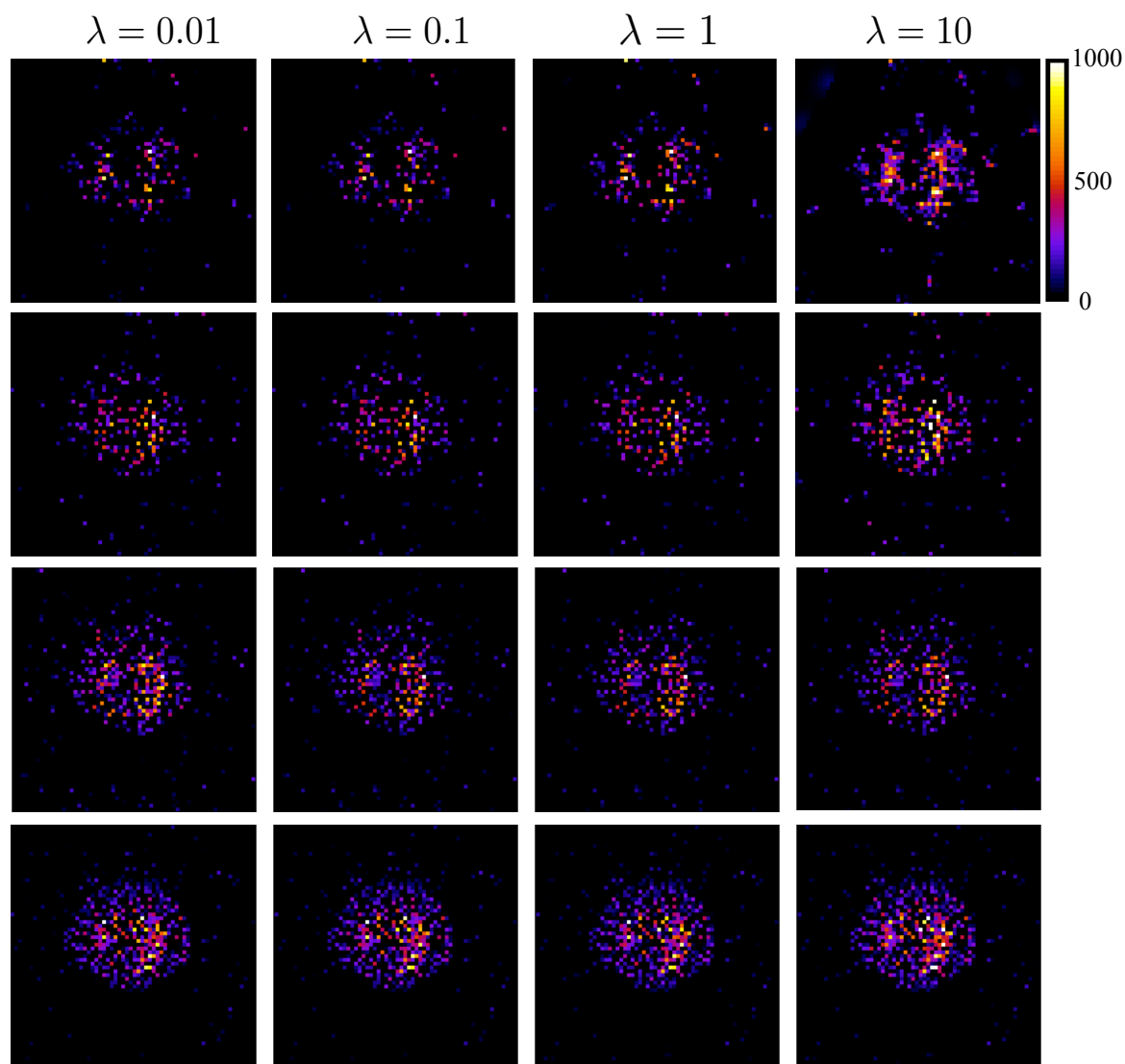
Figure 5.24: CP Reconstructions using the L_2 data constraint: each row represents a particular number of view angles and each column a given smoothing parameter λ . $s = 100$ in all cases.

voxels, even for very undersampled data, i.e., the streak artifacts were nearly eliminated using L_2 minimization.

The central slice intensity profiles shown in figure 5.25 reflect the noisy reconstructions showing large variations in intensity, and low SNR and CC compared to the profiles in figure 5.17 for TV-KL minimization.
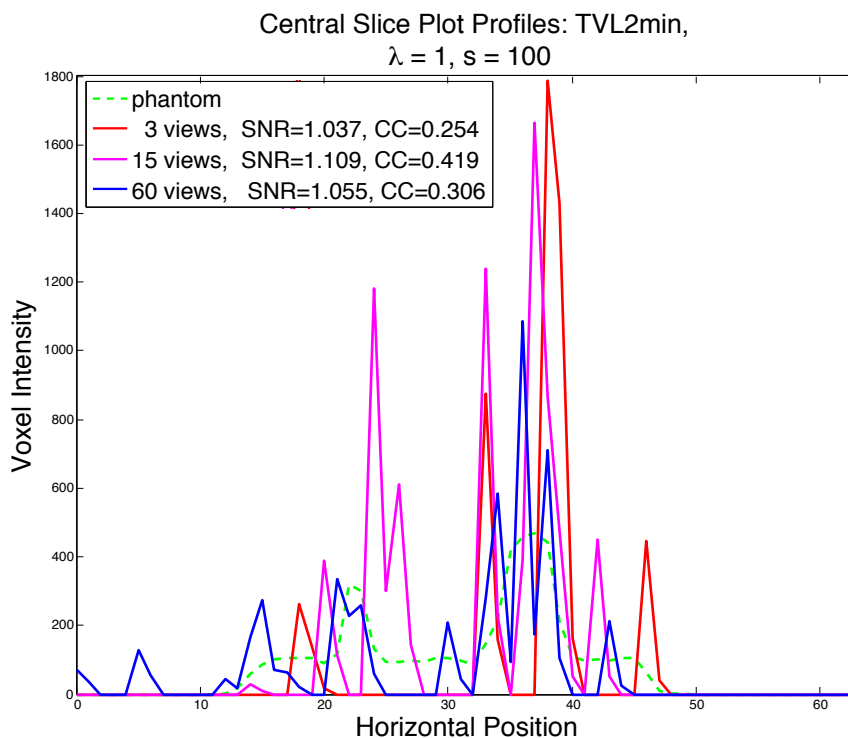
Figure 5.25: Horizontal intensity profiles for reconstructions of Poisson noisy data using the L_2 data constraint. The SNR and CC are calculated for the full reconstruction volume.

The estimated mean voxel intensity within the lung regions compared to the true mean intensity is shown in figure 5.26. Despite the noisy reconstructions, the mean estimates agreed well with the mean intensities of the phantom, and improved as the number of view angles was increased. These results are summarized in table 5.3, and are consistent with the results obtained using the D_{KL} data constraint, in that the slope was not affected by the choice of λ.
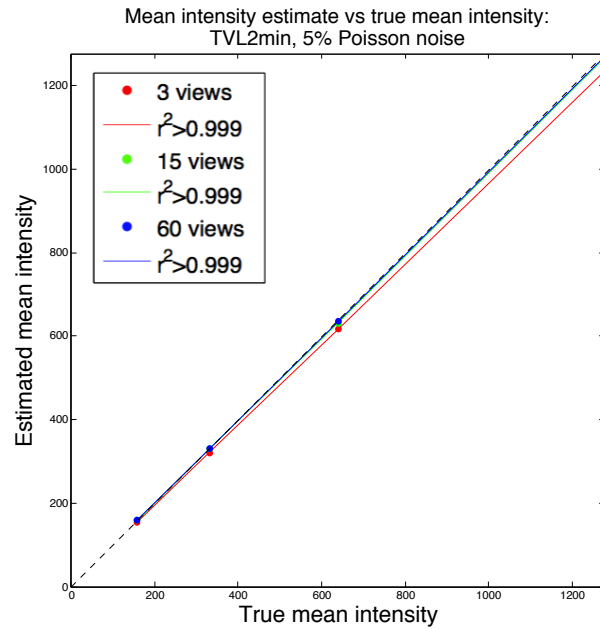
Figure 5.26: Mean intensity estimates in the lung regions for $\lambda = 1$ and $s = 100$ obtained using the CP method with the L_2 data constraint.

3 views	$\lambda = 0.01$	$\lambda = 0.1$	$\lambda = 1$	$\lambda = 10$
m	0.969	0.969	0.968	0.962
b	-1.313	-1.270	-1.133	0.181
r^2	>0.999	>0.999	>0.999	>0.999
15 views				
m	0.992	0.992	0.992	0.993
b	-0.456	-0.442	-0.414	-0.847
r^2	>0.999	>0.999	>0.999	>0.999
60 views				
m	0.993	0.993	0.994	0.993
b	-0.456	0.462	0.405	0.487
r^2	>0.999	>0.999	>0.999	>0.999

Table 5.3: Linear fits to the mean intensity estimates within the lung regions obtained using TV-L_2 minimization.

The convergence metrics using the L_2 data error in place of D_{KL} are plotted in figure 5.27 for $s = 100$. The L_2 data error decreased monotonically despite the theoretical discrepancy between the L_2 data metric and the Poisson noise added to the data. Generally, $PPDG \times n$ decreased with iteration number, but did not converge to zero. The
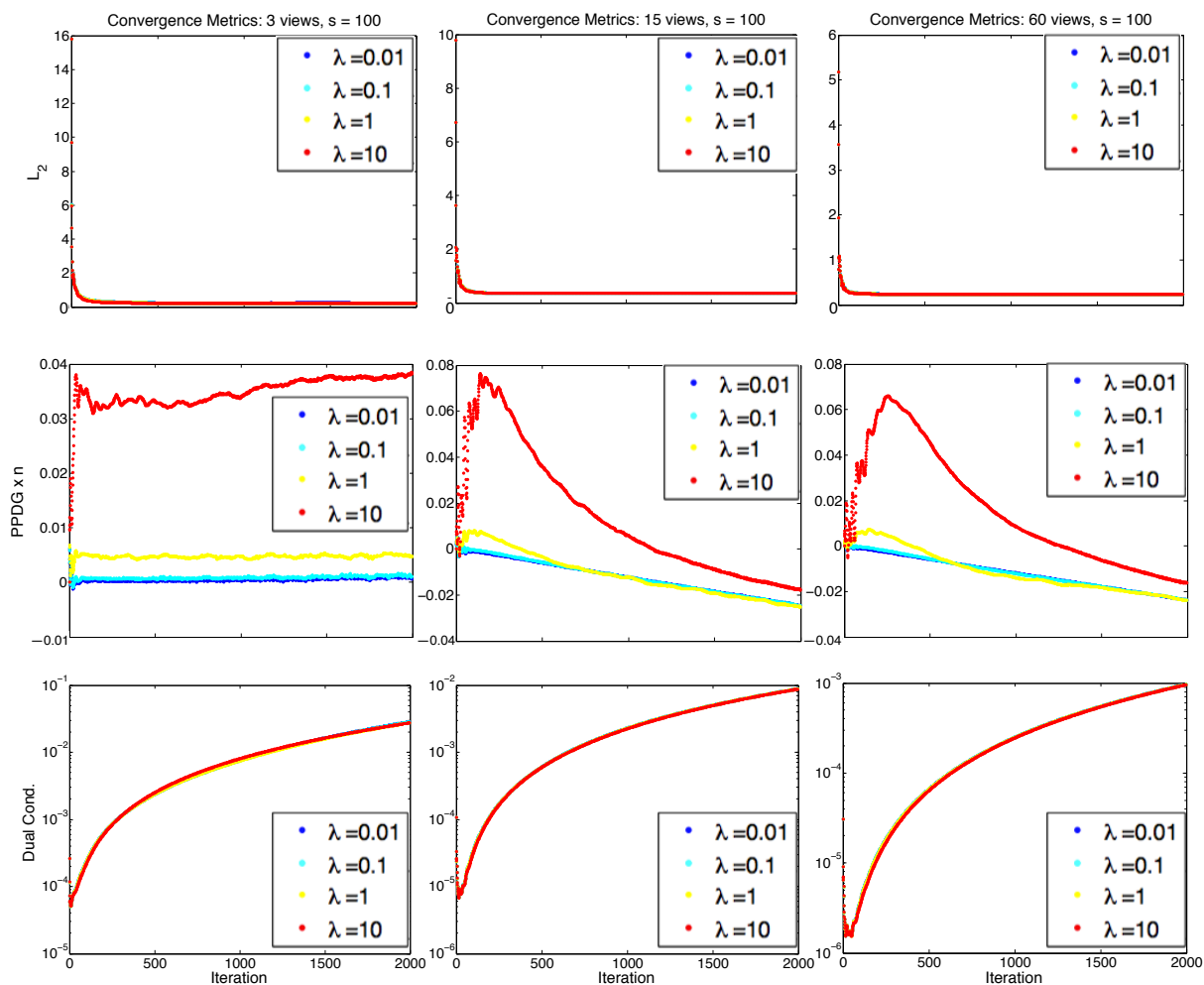
Figure 5.27: The CP convergence metrics: The rows correspond to L_2 data error, *PPDG* × *n*, and dual-condition. The columns display results for 3, 15, and 60 view angles. 5% noise.

dual-condition diverged, indicating that the method continued to perform significant updates to the reconstruction even after 2000 iterations.

The effect of the scaling parameter was investigated by reconstructing the lung phantoms with different mean activities using $\lambda = 1$ and letting $s = 1, 50,$ and 100. The results in figure 5.28 for 3 and 15 view angles show that the mean intensities agreed well with the known intensities for both cases. Changing *s* had no effect on the slope of the

mean intensity estimates, while increasing the number of view angles improved the mean

estimates, consistent with results obtained using the D_{KL} data error metric.
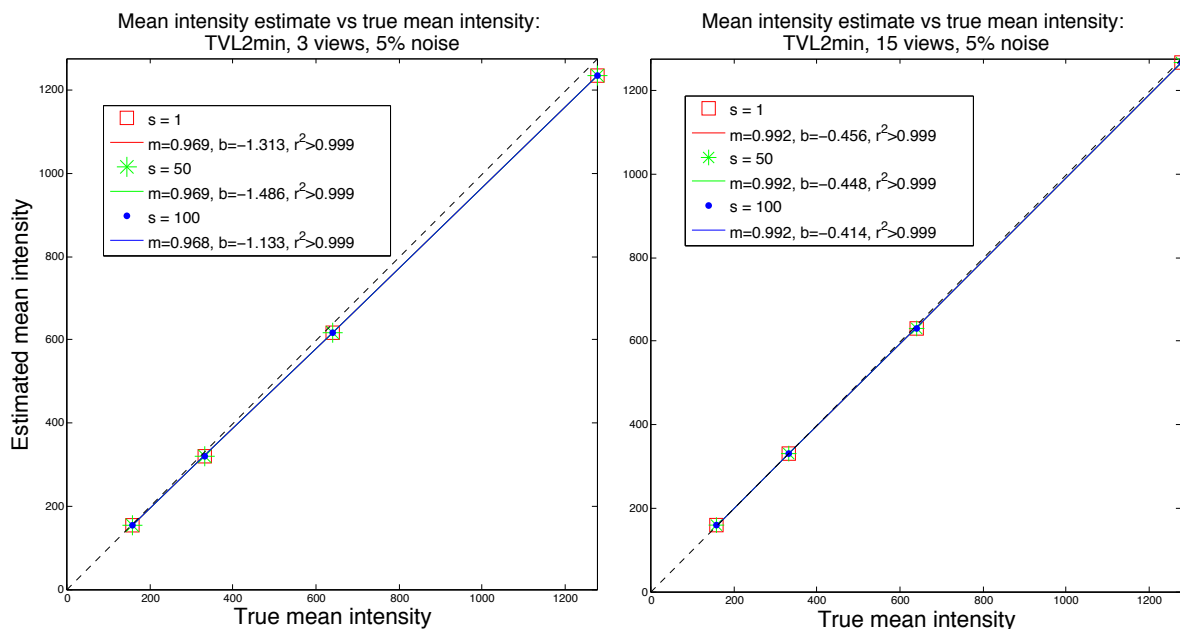
Figure 5.28: Mean intensity estimates in the lung regions for $\lambda = 1$ and $s = 1$, 50, and 100 obtained using the CP method with the L_2 data constraint.

5.4 Discussion

This chapter began by investigating our rotation-based matrix implementation of

the CP algorithm using a quadratic objective function applied to a simple phantom that

was sparse in gradient magnitude. Reconstructions using $\lambda = 1$ and 10^{-1} had a high degree

of agreement with the phantom, with the partial primal-dual gap converging faster than $1/$

n over 2000 iterations. The objective function was then replaced with that of the TV-KL

minimization problem and simulations with the simple phantom were repeated. Rapid

convergence was demonstrated for $\lambda = 10^{-4}$ and $\lambda = 10^{-3}$ and the simple phantoms were

recovered with a high degree of accuracy. Likewise, the CP method was robust enough to reconstruct phantoms that were not sparse in gradient magnitude, as demonstrated in section 5.2.5. *These results suggest that the forward- and back-projector operators are well matched and are appropriate for use with the CP method.*

Extending the rotation-based matrix method to our laboratory's system geometry involved incorporating the finite size of the multi-pinhole collimators into the forward- and back-projection operations. In this way, the CP method was tested by performing a more realistic SPECT reconstruction problem using a digital rat lung phantom with Poisson noise applied to the projection data. Our results showed that the smoothing parameter λ controlled the noise level in the reconstructions, consistent with its role as a noise penalty term in the objective function in equation 4.41. For a given s, small smoothing parameters gave relatively noisy reconstructions as measured by SNR, whereas large λ resulted in relatively blurry reconstructions. Nonetheless, when the mean intensity in a ROI of the reconstruction was calculated it was relatively independent of λ, i.e. even noisy reconstructions successfully achieved accurate voxel intensities, in an average sense. *This is important since one of the primary objectives of developing a reconstruction technique for few-view SPECT is to have a method that accurately recovers mean activity estimates for the quantification of radiopharmaceutical uptake in an organ/tissue of interest.*

The choice of λ has implications for other image quality features. When λ was small, the CP method resulted in more noise in the reconstructions, but also promoted sharper lung boundaries (edges). However, small λ resulted in more pronounced streak artifacts, likely due to the "sharpening" effect of small λ. Correspondingly, for large λ

(i.e., $\lambda = 10$) the edge features are blurred, leading to activity distributions that extended beyond the ROI and a reduction in the peak activity estimates within the ROI.

For a fixed number of view angles and fixed λ, our results show that the mean activity estimates within an ROI of the reconstruction is dependent on the choice of scaling parameter s. When s was small (i.e., $s = 1$), the CP method did not accurately estimate the mean intensities in all cases in figure 5.15. However, when $s = 50$ and $s = 100$, average intensity estimates in the lung ROI were more accurate and outperformed MLEM over the entire range of intensities for 3 and 15 view angles.

The choice in s also has important implications for image quality. Namely, when s is increased an overall reduction in streak artifacts around the periphery of the reconstructions was observed, particularly for 3 and 6 view angles in figure 5.20.

The CP method was also evaluated by determining the effect of λ on the convergence metrics introduced in chapter 4 and section 5.2.3. The *PPDG,* the dual-condition, and the KL data error, were calculated. The convergence metrics decayed to an asymptotic value for a given λ, suggesting the method reaches a point where the reconstruction ceases to improve with more iterations. As λ increased, this point was achieved sooner as seen in figure 5.18. However, reconstructions obtained using larger λ also demonstrate larger data error, larger *PPDG*, and larger average voxel updates after 2000 iterations. *Thus, as predicted, there is a tradeoff between data fidelity (KL data error) and smoothness, which is controlled by λ. Therefore, using a large smoothing constraint limits the CP method's ability to minimize the objective function.*

Likewise, the performance of the CP method was evaluated by determining the effect of s on the *PPDG*, the dual-condition, and KL data error. For a given λ, the

convergence metrics demonstrated smaller primal-dual gaps, smaller average voxel

updates, and better data fidelity as *s* increased (figure 5.23). However, increasing *s*

without bound can negatively affect the convergence of the *PPDG* and the dual-

condition. Specifically, the rate with which the *PPDG* approaches the plateau feature

observed in figures 5.18 and 5.23 can be positive. If convergence is characterized as the

PPDG monotonically decreasing as the method approaches its solution, this was not the

case when *s* was too large. This is demonstrated by plotting the change in *PPDG*, or Δ

PPDG × *n* over iteration in figure 5.29(a) and shows that for *s* > 500, the rate of change

is positive, so the *PPDG* is increasing as it approaches the solution. Likewise, the dual-

condition shown in figure 5.29(b) is not monotonically decreasing for *s* > 500. Both of

these violate the convergence criteria laid out in Sidky, et al [91, 92].

As the number of view angles used in the TV-KL reconstructions increased,

improved mean intensity estimates and better resolution of the boundary features in the
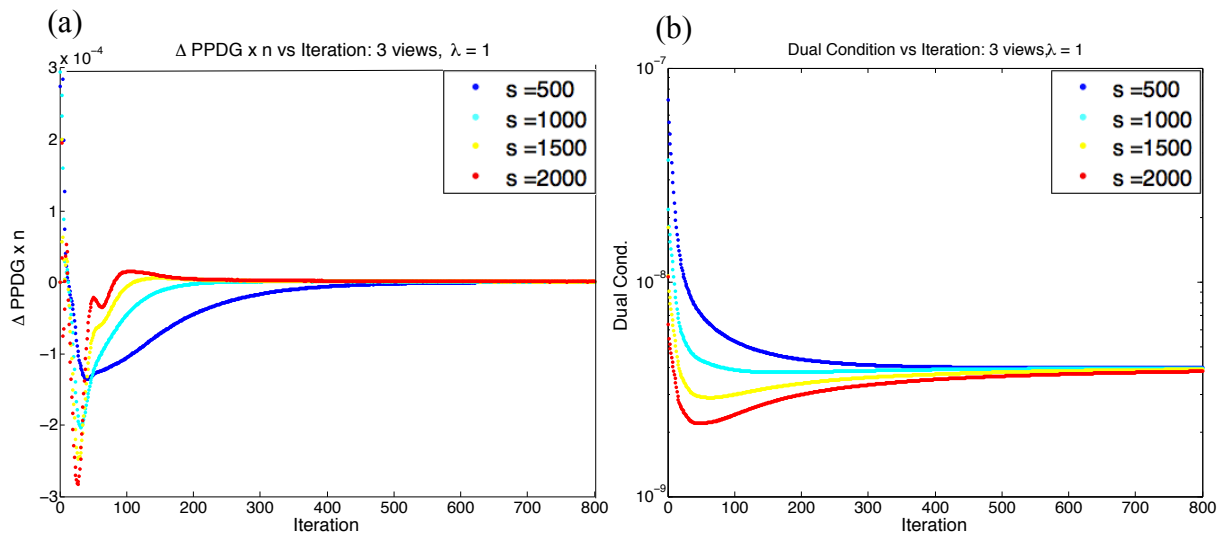
Figure 5.29: (a)Δ*PPDG* × *n*, the change in the partial primal-dual gap times the number of iterations as a function of iteration, (b) the dual-conditions plotted over iteration.

reconstruction were observed for a given choice of λ and s, as expected. However, in figure 5.16 and 5.17 we see that with only 3 view angles the CP method provided intensity distributions closer to the simulated lung distributions than MLEM.

Based on our implementation and evaluation of the CP method, there are practical and theoretical implications for applying TV-KL minimization to real data. Since the CP method generates accurate mean intensity estimates within the lung ROI when s is large, the method then favors large τ and small σ step-sizes, when $\sigma = 1/sL$ (the dual step size) and $\tau = s/L$ (the primal step size). Conceptually, if the proximal mappings are used as a generalized descent procedure, then optimal sizes for σ and τ reflect the relative topography of the solution space for the primal and dual problems. If the solution space of the primal problem is dispersed then minimizing the primal problem will take longer if the step-size τ is small. If σ and τ are inversely proportional, as in equation 4.43, this implies a very localized dual problem. For all smoothing parameters, increasing s improves data fidelity, decreases the primal-dual gap and the dual-condition. Additionally, based on the convergence metrics in figure 5.23, the solution obtained using the CP method is approached faster for $s = 100$ than $s = 1$, and suggests $s = 100$ is a more appropriate scaling between the proximal step-sizes for TV-KL minimization.

Unlike the ideal simulations, the *PPDG* did not decrease to zero faster than $1/n$ using TV-KL minimization in our lung phantom simulations, as shown in the left column of Figure 5.30. Rather, the rate of change of the *PPDG* decreases in magnitude faster than $1/n$. Plots of the change in *PPDG*, $\Delta PPDG \times n$ are given in the right column of figure 5.30. Using the CP algorithm for TV-KL minimization on the simulated lung problem, *PPDG* decays to a small value. Although slow convergence may be expected since TV-
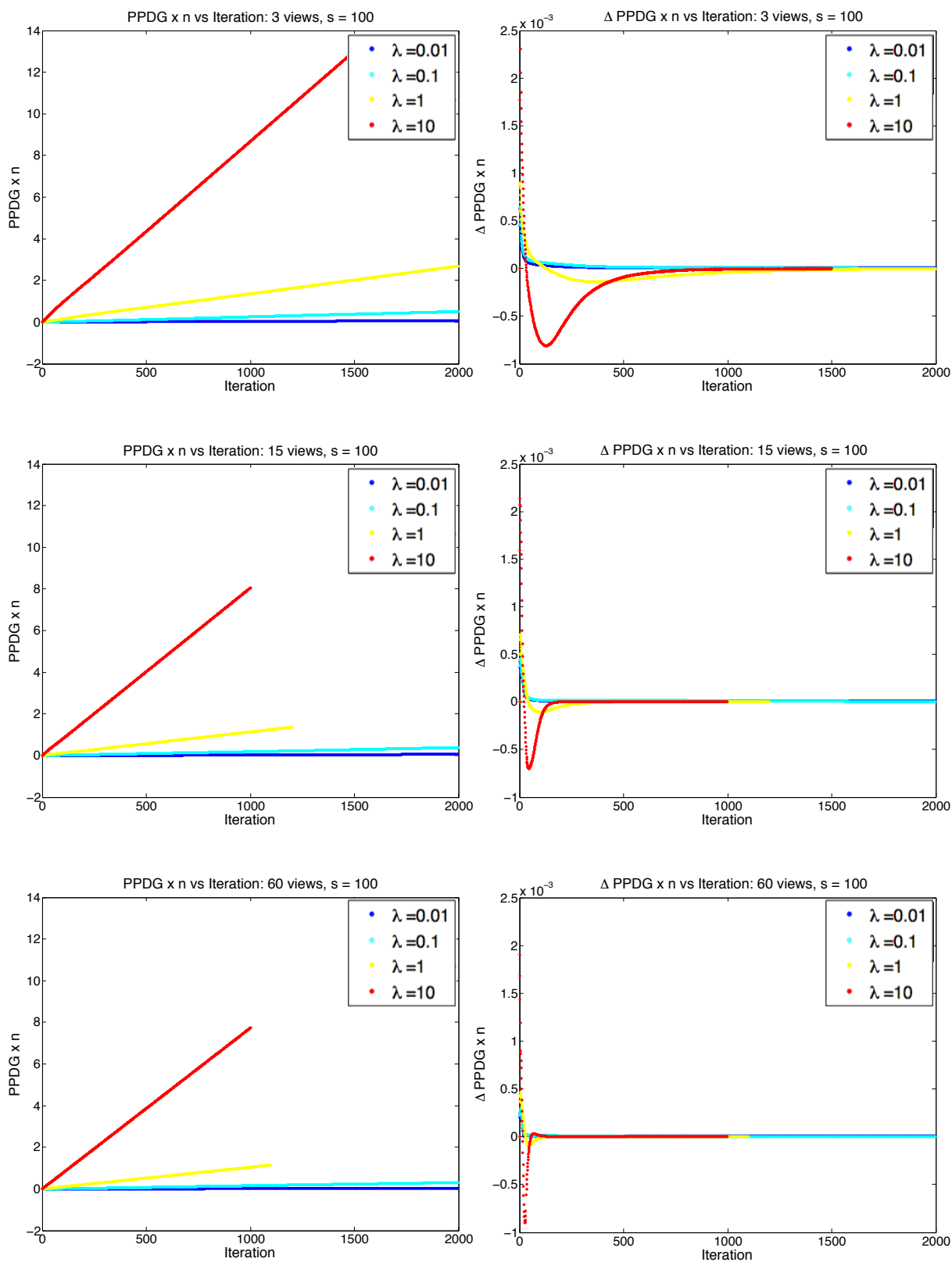
Figure 5.30: *PPDG × n* over iteration in the first column and Δ *PPDG × n* over iteration in the second column for 3, 15, and 60 view angles.

KL minimization is a constrained problem and is not quadratic, the decay feature of the

PPDG for the SPECT problem differs from the convergence behavior observed in the

ideal pinhole simulations in section 5.2.3. Since it would be impractical to iterate until the

PPDG is smaller than some small value, the size of the rate of change of the primal-dual

gap may be used to indicate the point at which a change in *PPDG* is unlikely to improve

without a significant number of iterations. Likewise, since the dual-condition is a

measure of the size of the update to the intermediate solution f_n, and since it decays to

some small value, the rate of change of the dual-condition may also be used as a stopping

criteria. One strategy for TV-KL minimization is to iterate until the rate of change of both

the primal-dual gap and the dual-condition are within some small number δ. In contrast,

the strategy presented here was to fix the total number of iterations to 2000 and use the

convergence criteria to judge whether the CP method converged. Since the number of

iterations required for convergence can vary by several orders of magnitude, it may be

impractical to wait until the dual condition or $\Delta PPDG \leq \delta$, particularly with the choice

in *s*.

Finally, the effect of using a different data error constraint on the reconstructions

was investigated using the TV penalty. Using the L_2 data error constraint to reconstruct

Poisson noisy data resulted in noisy reconstructions with little dependence on either λ or

s. Furthermore, based on the convergence metrics, the *PPDG* appeared to converge for

the smallest smoothing parameters, $\lambda = 0.01$ and $\lambda = 0.1$, which favor noisy

reconstructions. However, the dual-condition was not monotonically decreasing with

iteration regardless of λ, *s*, or view angle. This indicates more variation between

successive estimates after 2000 iterations compared to using the D_{KL} data error constraint.

Activity distributions that better fit the simulated distributions can be achieved when the data error constraint matches the noise model of the data. Despite the noisy reconstructions, using TV-L$_2$ minimization to reconstruct Poisson noisy data, the mean intensity estimates were close to the true mean intensity estimates within the lung ROI.

Chapter 6

Aim 3: Application To Experimental Data

Introduction

Dynamic 3-dimensional SPECT imaging requires imaging protocols and reconstruction algorithms that are tailored for providing functional or physiologic information. Thus, the objective of this section is to investigate the suitability of the CP method for reconstructing experimental data using TV-KL minimization. Experimental data likely contains systematic and random error due to an over simplified projection model resulting from subtle misalignments, differences in noise characteristics due to background or photon scatter, and imperfections in the camera, among other things. We performed two *in vivo* rat studies using TV-KL minimization to test the CP method over a range of activity levels. The first experiment involved multiple injections of a radiolabeled perfusion marker designed to asses the algorithm and its ability to accurately estimate mean activity in the lungs. The second experiment involved application of the CP method to a rat lung study using ^{99m}Tc-HMPAO.

Recent work [5, 6, 24] has demonstrated the use of ^{99m}Tc-HMPAO as an early biomarker for pulmonary oxidative stress using a single planar projection of the rat thorax. Oxidative stress refers to an imbalance between oxidants and anitoxidants, leading to a decreased ability of the lungs to detoxify reactive oxidant species [6]. These studies showed that lung uptake of HMPAO was increased in conditions in which oxidative stress was increased in the lung. This includes *hyperoxia*, prolonged exposure to a 100% oxygen environment as a model of ARDS, and lipopolysacharide (LPS)

administration, an animal model of sepsis. In the preliminary study here, we will

determine estimates of the lung uptake of ^{99m}Tc-HMPAO in control and hyperoxic rats

from fully 3D reconstructions that provide improved delineation of the lung region

compared to previous planar studies.

As demonstrated in the simulations, the CP method works well over a large range

in simulated activity levels. Reconstructions using large smoothing parameters λ generate

blurry solutions, potentially reducing the resolution of small features in the

reconstruction, while reconstructions using small smoothing parameters are sensitive to

noise. Results from the simulations in Chapter 5 suggest that the choice in λ has little

impact on the mean activity estimates in the lung regions. Improved estimates of the true

activity within a region of interest may be obtained by increasing the scaling parameter s.

The effect of changing λ and s on the experimental reconstructions will be examined

here.

6.1 Imaging hardware and geometry

Experiments were performed using the custom designed micro-SPECT system

discussed in chapter 2. A schematic of the overall experimental setup is shown in figure

6.1. The system consists of 3 stationary modular gamma cameras spaced 120° apart and

placed around a small animal specimen holder. The SPECT system uses five-pinhole lead

collimators as pictured in figure 2.19, where the diameter of each pinhole is 2 mm. The

collimators were inserted into windows cut into the sides of the triangular lead specimen

holder, which can accommodate collimators with different pinhole configurations. The

dimensions and geometry of the pinhole (diameter and angle of acceptance) are used in the subsequent calculations of the projection model.
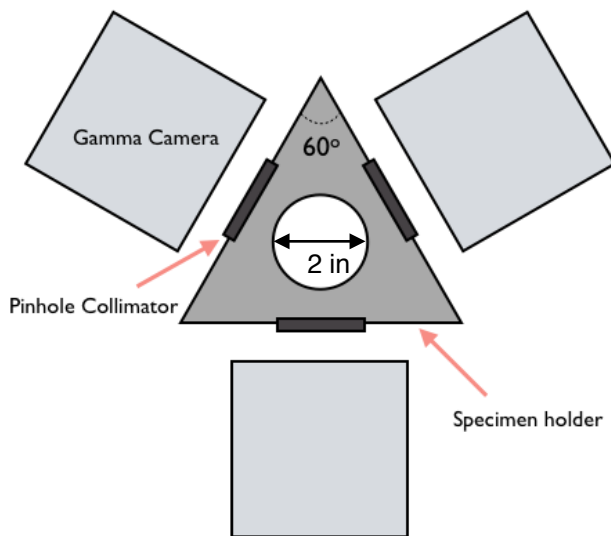
Figure 6.1: Schematic of the micro-SPECT system.

6.2 Experimental Setup

The schematic in figure 6.2 depicts the experimental setup that was used for the two imaging studies. Sprague-Dawley rats were anesthetized (50 mg/kg of Pentobarbital) and a 25 gauge, 3/4 inch catheter was inserted into the tail vein. The animal was placed in a vinyl tube with a 2 in diameter, which was then positioned vertically on the top of a rotation stage. To deliver the SPECT radiopharmaceutical, a thin cannula was connected to the tail vein catheter. The cannula was threaded through a hole in the bottom of the vinyl tubing, and connected to a three-way stopcock located behind the lead brick shown in figure 6.2(a) and (b). The lead brick served to shield the gamma camera while the injection was made. The cannula was 24 inches long so that it could easily wrap 360° around the rotation stage while connected to the stopcock assembly. The

radiopharmaceutical was injected through the top port of the stopcock. After each

injection the cannula line was flushed with saline-diluted heparin through the side port of

the stopcock. Although the specimen holder is designed for 3 gamma cameras, for this

project multiple projections were taken by rotating the animal in the specimen holder and

acquiring data using a single camera/collimator, as proof of principle. In this way,

misalignments between multiple view angles could not affect the reconstructions and

subsequent analysis. Supplemental anesthetic was given as needed.
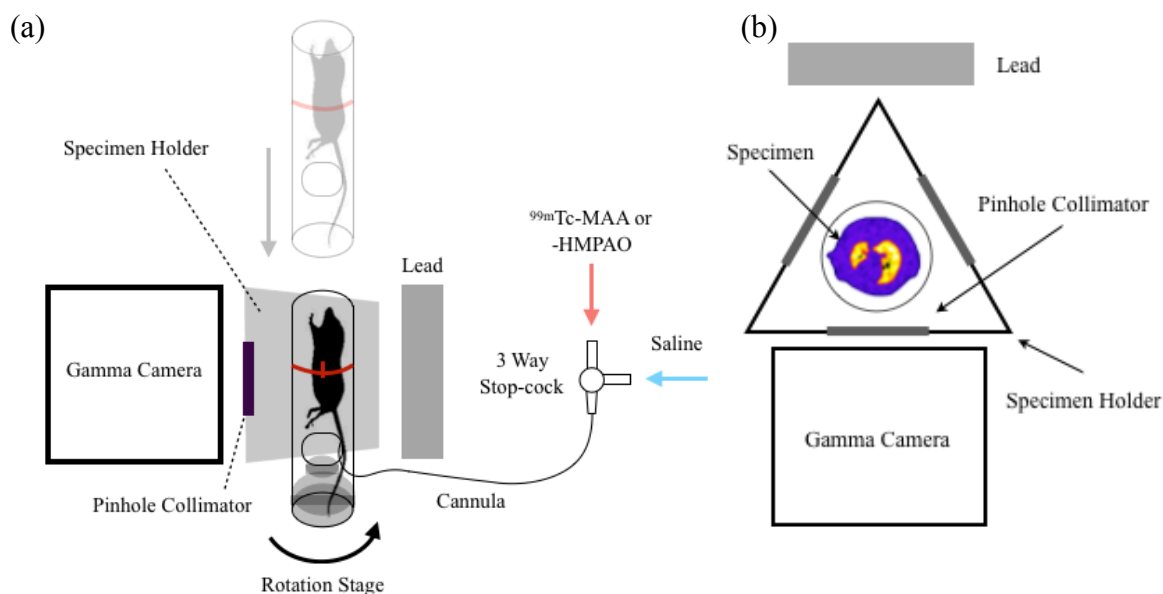
Figure 6.2: Schematic of the experimental setup: (a) side view [105], and (b) top-down view.

6.2.2 ^{99m}Tc-MAA imaging

Experiments were performed to determine the optimal design parameters of the

CP algorithm. For this experiment, technetium-labeled macroaggregated albumin (^{99m}Tc-

MAA), obtained from Cardinal Health, was used. An injection of ^{99m}Tc-MAA (0.2 mCi in

~0.2 ml) was made via the tail-vein. As described in chapter 2, >98% of [99mTc]-MAA gets trapped in the lung microvasculature, due to its size. The breakup and elimination of [99mTc]-MAA from the lungs has a half life of 2-3 hours [30], hence the distribution of [99mTc]-MAA of activity should not change over the study duration. After 5 minutes the rat was imaged using 60 equally spaced view angles over 360° with a scan time of 5 seconds per view angle. The [99mTc]-MAA injection and imaging was repeated using 0.15 mCi, 0.17 mCi, 1.57 mCi, 1.36 mCi, and 1.28 mCi, each in ~0.2 ml, for a total of six injections with a final accumulated dose of 4.73 mCi. The three lowest doses correspond to expected activity levels in the lung when performing HMPAO experiments. After each injection, an independent set of data was acquired with increasing activity, and reconstructions of each data set were performed. The 60-view projection data set was subsampled into data sets containing 3, 6, and 15 equally spaced view angles and the reconstructions were repeated for each case.

6.2.3 [99mTc]-HMPAO imaging

For these studies, the setup and protocol described above for delivering the radiopharmaceutical was the same as the [99mTc]-MAA rat experiments. Exametazime HMPAO (*CERETEC ®*) kits were purchased from GE Healthcare [16] and HMPAO was reconstituted and radiolabeled according to the kit instructions. An injection of [99mTc]-HMPAO (5 mCi in 0.8 ml) was made, and after 15 min steady state uptake in the lungs was reached [24]. The animals were then imaged over 60 equally spaced view angles using 10 second acquisitions. As observed in previous planar studies, [99mTc]-HMPAO is taken up by other organs and surrounding tissue, besides the lung, making it difficult to

identify the lung boundaries in the HMPAO reconstructions. Thus for each experiment a second injection of ^{99m}Tc-MAA (3 mCi in 0.2 ml) was made, and images were acquired for 5 seconds over 60 equally spaced view angles. The ^{99m}Tc-MAA data then contains activity from both the ^{99m}Tc-HMPAO and ^{99m}Tc-MAA injections. However, since the ^{99m}Tc-MAA is trapped within the lung vasculature, its contribution to the signal in the lung is dominant, so that these reconstructions can be used to identify the lung regions in the ^{99m}Tc-HMPAO images. We shall refer to these data as simply MAA and HMPAO images/reconstructions.

HMPAO data was reconstructed with the CP method and MLEM, and repeated using 3, 6, and 15-view subsampled data. The 60-view MAA data was reconstructed using 30 iterations of MLEM. The lung regions were identified in the MAA reconstructions using a threshold segmentation procedure. An intensity threshold was calculated using Otsu's method [38] on the histogram of the central slice of the reconstruction volume. Otsu's method calculates a statistical distance, or "level," between two dominant distributions from the grayscale histogram of an image. As illustrated in figure 6.3, the image histogram features a cluster of voxel intensity around zero, regions outside the rat lung boundary, and a smaller number of voxels with a wide range in intensity from the lung region. Thus, intensity values above the Otsu threshold are attributed to the lung region. This threshold was then applied to the whole reconstruction volume, to identify voxels with intensities greater than the threshold level resulting in a mask of the lung region. The mean HMPAO activity per voxel within the lung region was then calculated based in the MAA mask.
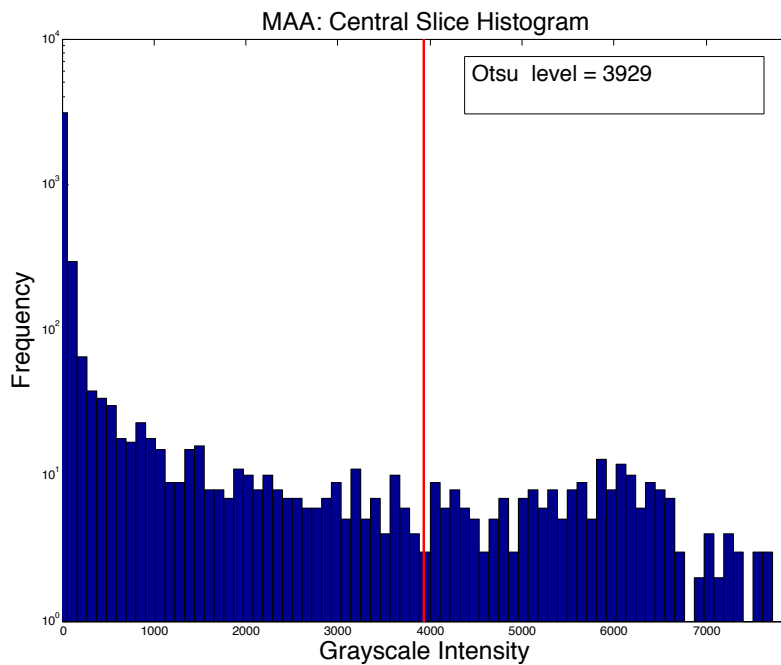
Figure 6.3: Grayscale histogram obtained from the central slice of an MAA reconstruction.

6.3 Considerations For Reconstructing Experimental Data

Since the forward operation in the case of *in vivo* studies corresponds to experimental acquisition of SPECT projections, it is important that the forward-projection operation accurately reflects the geometry of the experimental setup, regardless of the reconstruction method. This is particularly important for our implementation of the CP reconstruction algorithm, since the forward- and back-projection operators must be well matched, as discussed in chapter 4 and appendix C. Even small errors in the forward model of the system can result in artifacts in the reconstructions, which in the case of CT most commonly manifest themselves as ring artifacts [106, 107]. Thus, we performed a series of procedures to precisely identify the

parameters of the forward model, including those related to the system geometry and the pinhole response.

The first task was to establish the parameters of the imaging system matrix **H**. To determine the distance between the effective image plane and the pinhole, a line source consisting of a glass capillary tube containing 1 mCi of Pertechnetate, TcO_4^-, in ~ 0.1 ml was positioned at the center of rotation of the SPECT system and projection data was acquired, as shown in figure 6.4. Using the line projection from each pinhole, an estimate of the image-to-pinhole (*i2p*) distance was triangulated from the projection image from each pinhole. For our system, the pinhole-to-center of rotation (*p2c*) distance is fixed at
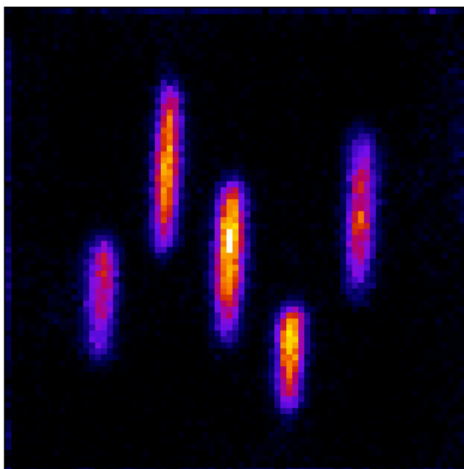
Figure 6.4: Planar projection of a capillary tube, containing 1 mCi of activity, placed at the center of rotation of the imaging system and imaged with the multi-pinhole collimator.

5.3 cm. The estimate of *i2p* was further refined by performing a phantom experiment. The phantom pictured in figure 6.5 was constructed, consisting of 7 tubes each with a diameter of 6.5 mm spaced 19 mm radially from the center of the central tube. Each tube was filled with 0.6 mCi of TcO_4^- mixed with saline in 0.8 ml. The phantom was imaged

(a)   (b)
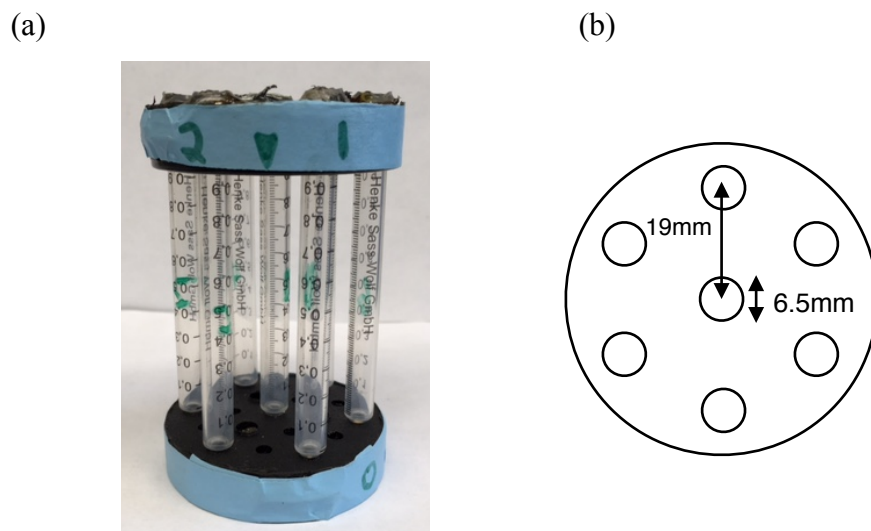
Figure 6.5: The 7 tube phantom: (a) Phantom, (b) top-down schematic depicting the tube
diameters and radial spacing.

using 60-views and MLEM reconstructions were obtained using the initial estimate of *i2p*

calculated from the line source. MLEM reconstructions were then repeated on the same

data using a series of different *i2p* values. Figure 6.6 shows MLEM reconstructions

obtained by perturbing *i2p* in 2 mm increments from the initial estimate from the line

source. The optimal *i2p* distance was found to be 3.2 cm. The system matrix **H** was

constructed with these parameters; ray tracing was performed using the geometry of the

(a)   (b)   (c)   (d)
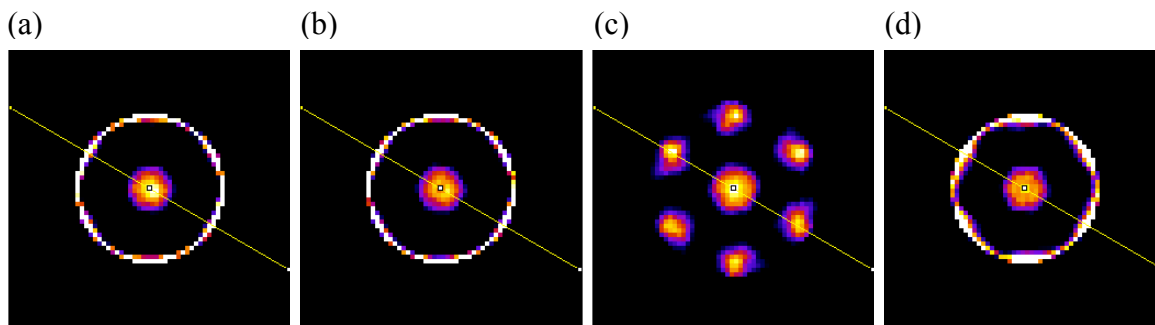
Figure 6.6: MLEM reconstructions of the activity phantom, used to find the image-to-
pinhole-distance (*i2p*): (a) *i2p* = 2.8cm, (b) *i2p* = 3.0cm, (c) *i2p* = 3.2cm, and (d)
i2p = 3.4cm.

pinhole collimator, *i2p,* and *p2c* to calculate the geometric weights as described in

chapter 2.

6.3.1 Alignment to the center-of-rotation

An important task prior to any tomographic study is to determine the location of

the center-of-rotation on the projection plane. This was done by imaging an extended

object over many view angles. A capillary tube containing 1 mCi of Pertechnetate was

placed in front of the camera at an angle. Thirty-five equally spaced projections were

acquired over 360°. Figure 6.7(a) shows a projection of the capillary tube. The sinogram

[53] of the central slice generated by plotting the intensity of each view angle, θ, as in

figure 6.7(b), where the intensity of each pinhole projection traced out over θ looks

characteristically like a sine curve. The center of rotation in the *x*-direction in figure

6.7(b) was located by finding the centroid of the sine-like curve attributed to the central

(a)   (b)
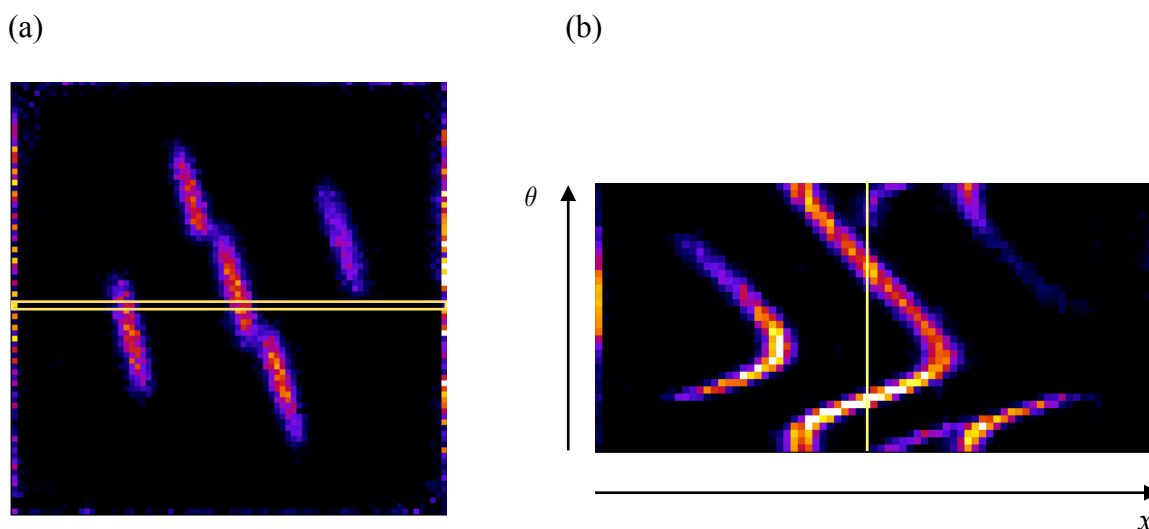
Figure 6.7: (a) projection of a capillary tube filled with 1 mCi of Pertechnetate positioned at an angle, (b) the sinogram (or Radon Transform [53]) generated using a single slice through the projections over all view angles.

pinhole. Each projection was then shifted so that the center-of-rotation lined up with the center of the *x*-direction.

6.3.2 Image preprocessing

In order to better match the forward model to the data, detections that are not consistent with the ray tracing model must be removed to avoid introducing artifacts into the reconstruction. Most background activity in the projection data is due to residual activity in the inlet cannula that wraps around the rotation stage in figure 6.2. Although lead shielding was added to mitigate detections from underneath the specimen stage, due to spatial limitations, some activity is detected at the bottom of the projections in figure 6.8(a) that cannot have passed through pinhole openings.

(a)   (b)   (c)
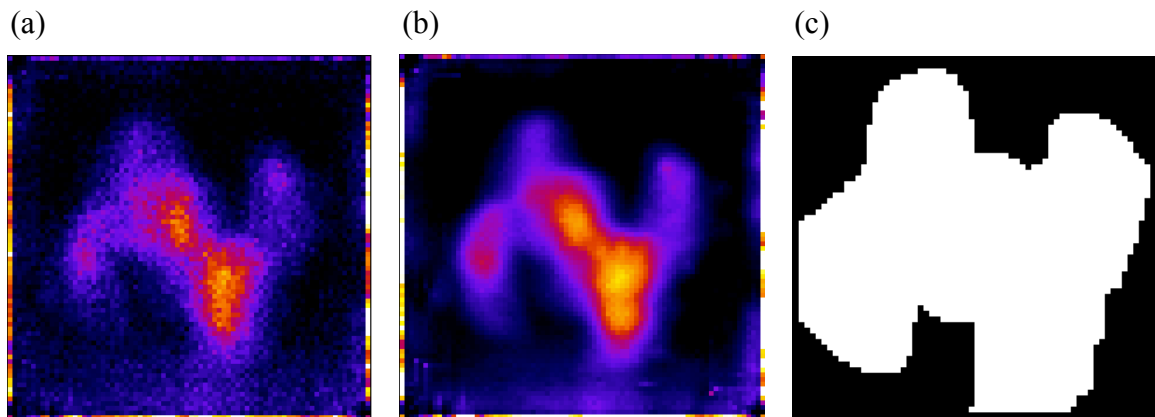
Figure 6.8: (a) Raw HMPAO data, (b) sum of all projections, and (c) mask of the pinhole projections obtained using the sum projection from multiple datasets.

To eliminate these detections, the projections from all view angles were summed and used to create a mask of the pinhole. By combining the sum projection from multiple view angles, activity is summed across the solid angle of acceptance of each pinhole.

This effectively creates a projection of a flood onto the detector plane, where the individual pinhole projections could be easily masked out using a threshold procedure. Furthermore, by combining the sum projections across multiple acquisitions (both MAA and HMPAO) the flood mask excludes pixels from the projections that cannot be traced through the collimator and the animal holder.

Another potential source of error may be corrected by restricting the relatively large angle of acceptance of the pinholes. The solid angle acceptance allows for the possibility of photon detections along rays traced outside of the physical location of the animal holder. Thus, the large angle of acceptance can result in data being back-projected off into regions outside of the animal. This can result in streak artifacts due to undersampling of the data. In order to maximize the amount of data back-projected into the region where there is activity, the diameter of the animal holder was used as an additional constraint so that only rays that trace through the diameter of the animal holder were used to to calculate the system matrix $\mathbf{H}$. This results in a back-projection operation in which data converges to the animal, rather than diverging to the periphery of the reconstruction.

Additional freedom was incorporated into the ray-tracing geometry used to generate the system matrix, using the center position of the large bore diameter of each pinhole. A mask similar to figure 6.8(c) was generated by forward-projecting a digital flood volume. The forward-projection model was aligned with the experimental flood projection by digitally adjusting the center location of both the front and back of each individual pinhole until the two masks overlapped. Figure 6.9 shows the central slice of

an ^{99m}Tc-MAA reconstruction with and without aligning the flood masks, demonstrating a reduction of ring artifacts.

Finally, artifacts around the periphery of the projections, due to the ML position estimation used in the cameras [94] and potential reflections off the edge of the light guide, can be seen in figures 6.7 and 6.8. To correct for this, each projection is cropped from a 78×78 grid down to a 64×64 grid prior to reconstructing the data.

(a)   (b)
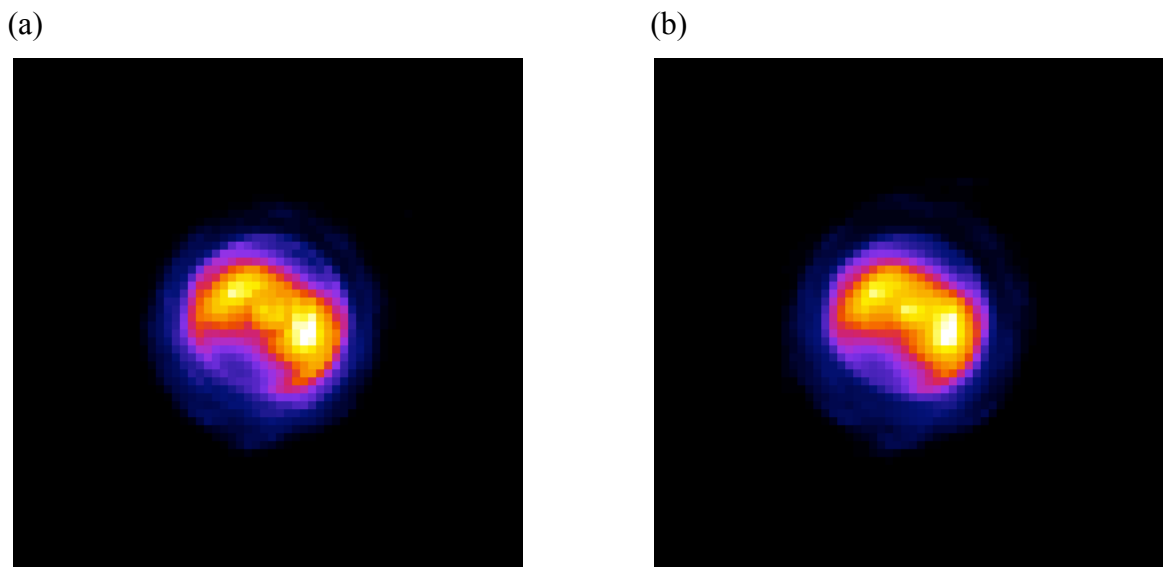
Figure 6.9:^{99m}Tc-MAA reconstruction using MLEM and 60 equally spaced view angles; (a) before adjusting the pinhole positions, and (b) after digitally adjusting the pinhole positions.

6.3.3 Normalization of the ray-driven multi-pinhole system matrix and ring artifact correction

The ray driven projection model used to calculate the system matrix has been described in the literature by Wietholt [102] for the case of single-pinhole SPECT, and is normalized according to the procedure outlined in chapter 2, where the columns of the

projection matrix are normalized to 1. In the case of multi-pinhole collimation, normalization of the system matrix may be carried out in two different ways; (1) sum the geometric weights of all rays passing through all of the pinholes of the collimator and normalize as if one view angle is one projection, or (2) normalize the geometric weights of the rays from each pinhole separately, sum the results, then divide by the number of pinholes. Because the simulations use the same forward-projector to generate pseudo-data, either normalization procedure may be used in the reconstruction of phantoms. To determine which strategy better matched the experimental forward model, a rat ^{99m}Tc-MAA dataset was reconstructed using MLEM and both system matrices. Figure 6.10(b) and (c) show the pseudo projections obtained after 30 iterations using procedure (1) and (2), respectively. Normalizing the the system matrix according to procedure (1) resulted in pseudo-data with intensities that more closely match the projection intensities in the experimental data, suggesting its use in our algorithm. Appendix D suggests an additional modification that can be made to procedure (2) to better match the experimental

(a)   (b)   (c)
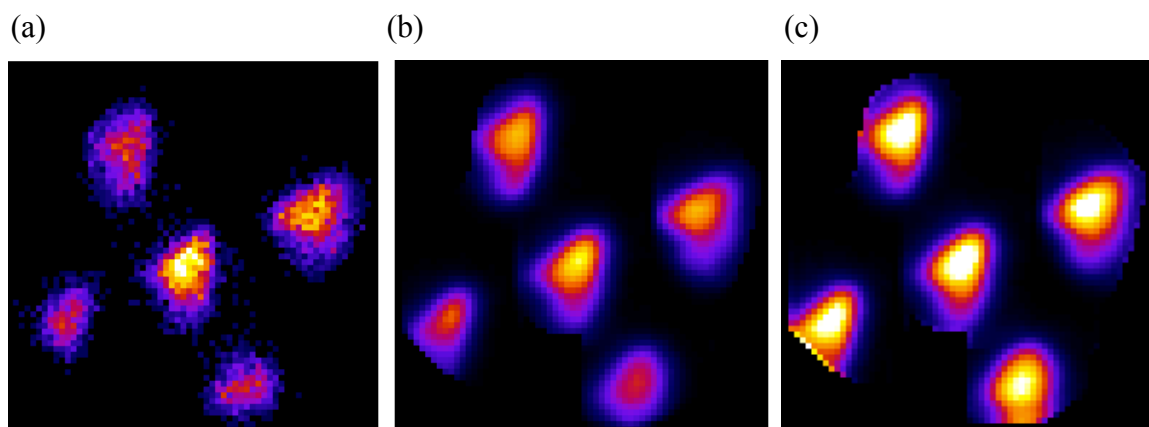
Figure 6.10: Normalization strategies: (a) experimental ^{99m}Tc-MAA projection, (b) forward-projection using procedure (1), (c) forward-projection using procedure (2).

projections, however it requires manually weighting the response for each pinhole for a given set of data, and is less adaptive.

Reconstructions obtained using 60-view angles are shown in figure 6.11, where the image on the left uses the normalization procedure (1) and the image on the right uses procedure (2). Ring artifacts were more apparent when treating a single view angle as one projection, using procedure (1). The ring artifacts were mostly eliminated, particularly in the center of the reconstruction, using the matrix generated from procedure (2). These

(a)   (b)
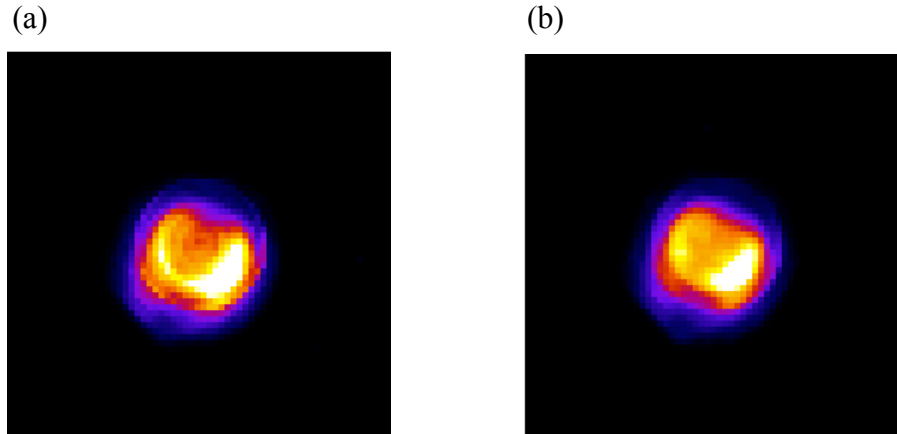
Figure 6.11: Normalization strategies: (a) normalizing using procedure (1), (b) normalizing using procedure (2).

artifacts may be attributed to slight variations in the back-projector's response due to transitions in angular coverage and the number of rays used in the ray tracing model.

The final improvement incorporates the response of the back-projector to a flood projection. Images in figures 6.12 were obtained by back-projecting a flood projection using normalization strategy (1) for 1 and 60-views angle ((a) and (b)), and strategy (2) for 1 and 60-view angles ((c) and (d)). As evident in figure 6.12(b) and (d), more weight is given to regions of the reconstruction within a small circle from the center, which was

found to be due to the amount of coverage each voxel gets in the reconstruction volume. Voxels that are visible from all view angles through all pinholes have the most amount of coverage and are located in the center of the reconstruction volume. These voxels are the columns of the system matrix that normalize to 1. By back-projecting a dataset of ones (a flood) we can map the matrix normalization onto the reconstruction volume to see the "response" of the back-projector for a given system geometry. White represents 1 and black represents 0. Since the response is not the same for all voxels, the activity estimates

(a)   (c)

(b)   (d)
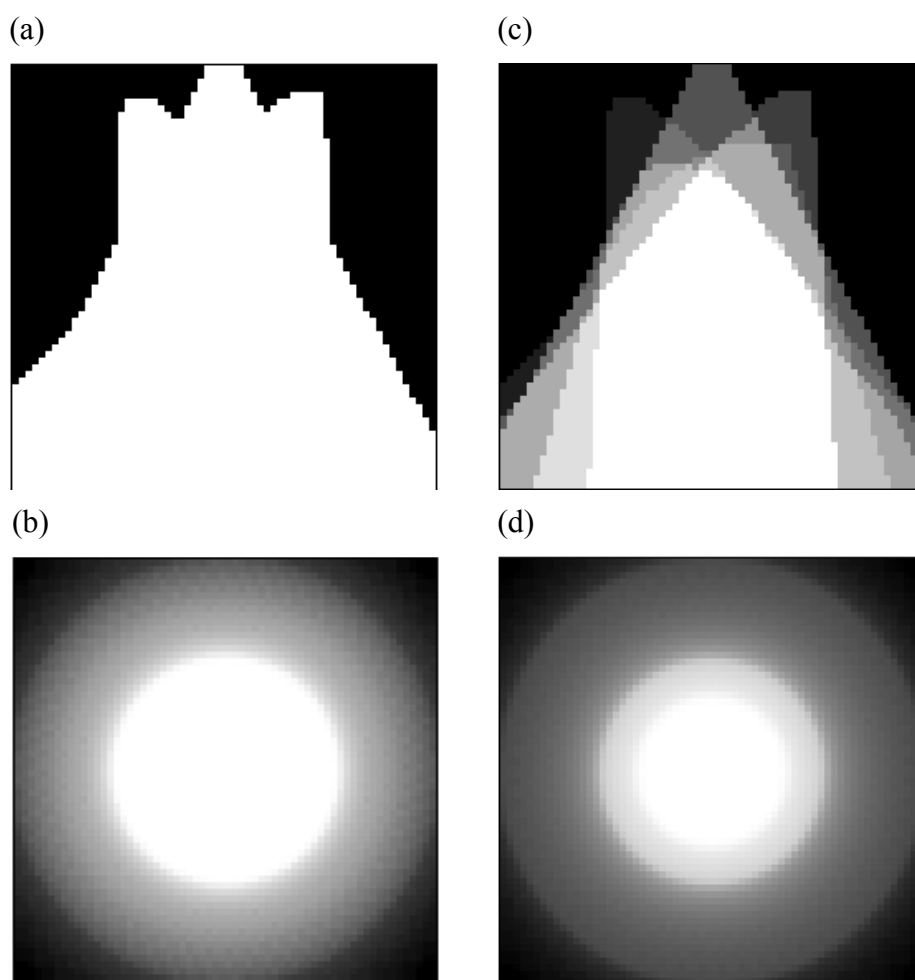
Figure 6.12: Back-projection of 1's; (a) one view angle using procedure (1), and (b) using 60-view angles. (c) one view angle using normalization (2), and (d) using 60-view angles.

from voxels that fall outside the small circle of 1's in the normalization map will be less than those from within the small circle for the same amount of activity. This will result in errors in quantification if the distribution of emissions extends beyond this small circle.

Experimentally, this problem could be corrected by increasing the pinhole-to-center of rotation distance or the image-to-pinhole distances, however both of these are fixed in our system. Alternatively, using the system matrix approach, the back-projector's response was accounted for by finding the difference between a voxel's value in the normalization map and its ideal value of 1. This correction can even account for very subtle variations in the back-projector's response which can be caused by the ray tracing method. This difference is the amount by which the jth column in the system matrix needs to be corrected,

$$h_{ij} = h_{ij} \cdot \left(1 - w \cdot (\mathbf{H}^T \mathbf{1})_j\right) \tag{6.1}$$

where h_{ij} is contribution of the jth voxel to the ith pixel, from chapter 2. The weight, w, was fixed to 1. Using the normalization procedure (1), figure 6.13 shows a resulting slice
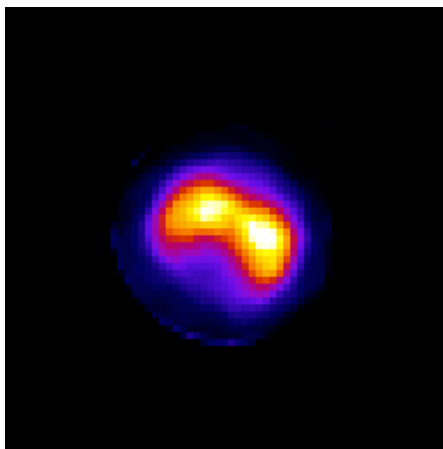
Figure 6.13: ^{99m}Tc-MAA reconstruction using MLEM and 60-view angles. The ring artifacts are largely eliminated after adjusting the pinhole positions and correcting for the back-projector's response to the imaging geometry.

of an MAA reconstruction using MLEM including all the adjustments incorporated to correct for misalignments and the back-projector's response in regions with limited coverage.

6.4 ^{99m}Tc-MAA activity estimates

This section will examine the CP reconstructions over a range of activity levels for *in vivo* rat data obtained using the multi-dose ^{99m}Tc-MAA imaging protocol. The effect of changing the smoothing parameter, the scaling parameter, and the activity, will be investigated, similar to the lung phantom investigation.

6.4.1 The effect of changing λ and number of view angles

We investigated the effect of changing the smoothing parameter on the multi-dose MAA data. The central slice of the reconstructions obtained using 3, 6, 15, and 60-views are shown in figure 6.14, with *s* fixed at 100. Each reconstruction was performed using TV-KL minimization for 2000 iterations, with the smoothing parameter varied between 0.5 and 5. The example reconstructions shown in figure 6.14 were generated from the 3.5 mCi MAA data set. MLEM reconstructions using 30 iterations are shown for comparison.

As before, increasing λ resulted in distributions that appear blurry and for $\lambda > 3$ the outlines of the lungs are obscured. Streak artifacts were present outside of the lung regions, but were reduced as the number of view angles was increased. The MLEM reconstructions were noisy, in comparison, with the exception of 60-view angles.

The central profiles for the $\lambda = 1$ reconstructions from figure 6.14 are plotted in figure 6.15. Behavior similar to the simulations in chapter 5 is observed; smoother
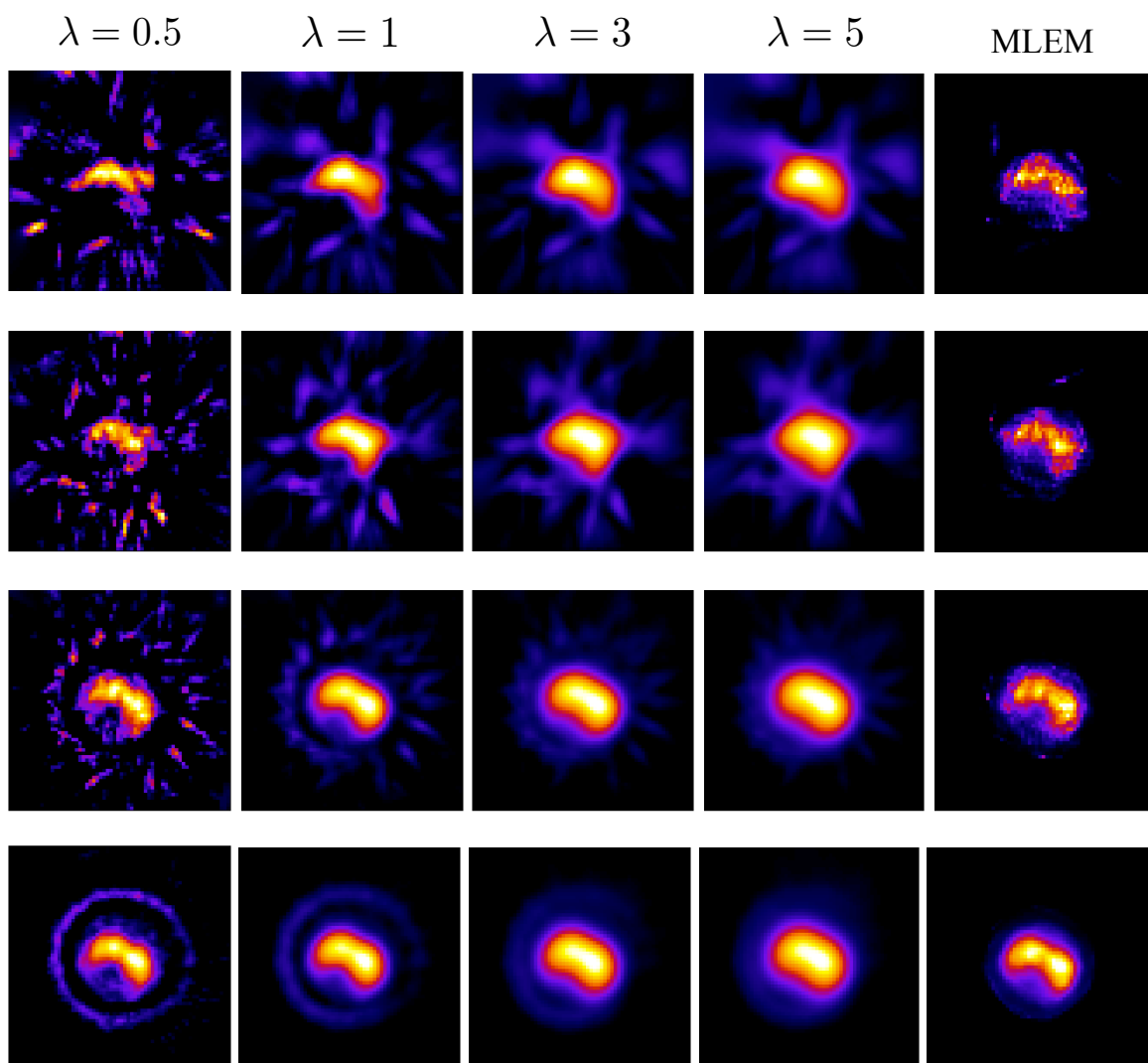
Figure 6.14: Reconstructions of 3.45 mCi ^{99m}Tc-MAA distribution. Each row represents reconstructions with a fixed number of views, and each column represents reconstructions with a fixed smoothing parameter.

intensity profiles were achieved, but the mean of the profile was constant as the number of views increased. Also, more activity was present outside of the lung region for 3-view angles and the peak intensity was lower. For MLEM the profiles in figure 6.15(b) were noisier than those from the CP method, but did not contain noticeable activity outside the
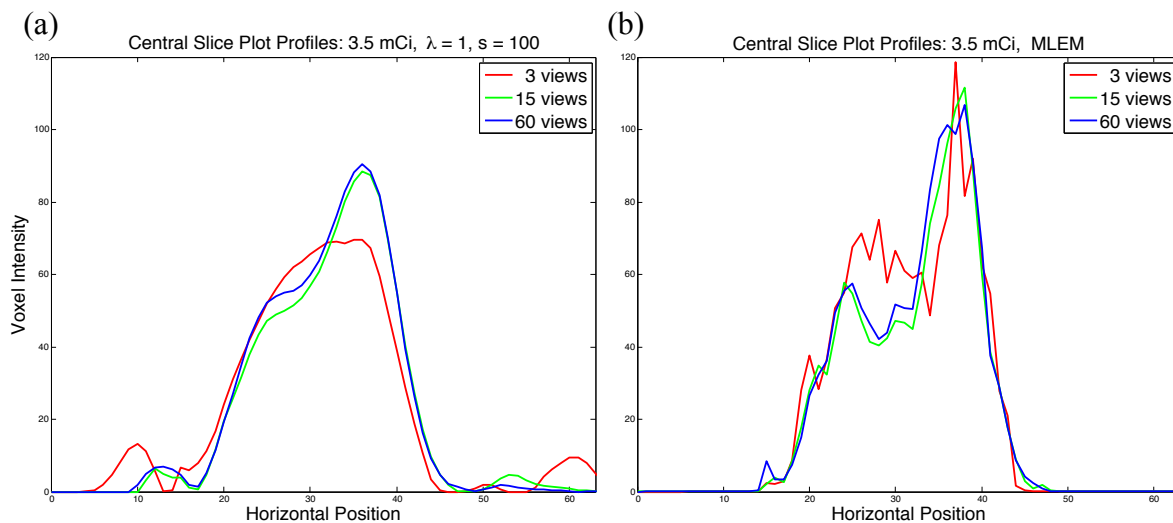
Figure 6.15: Horizontal intensity profiles taken from the central slice of the 3.5 mCi data using 3, 15, and 60-view angles: (a) the CP method using $\lambda = 1$ and $s = 100$, (b) MLEM.

lung boundary. This suggests that MLEM was less sensitive to variations in the back-projectors response.

To obtain the mean intensity estimates within the lung regions, a lung mask was created using the 60-view CP reconstruction of the 4.73 mCi data using $\lambda = 1$. As illustrated in figure 6.16, the lungs were segmented using Otsu's method, a binary mask was created, and the mean value within the lung volume was calculated for each
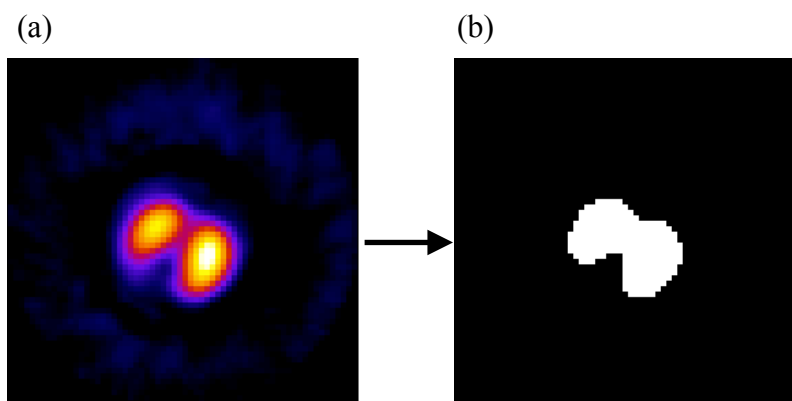
Figure 6.16: (a) The 4.73 mCi reconstruction, (b) the lung mask after thresholding the reconstruction.

reconstruction. Figure 6.17(a) shows a plot of the mean activity estimates using the CP method for 3, 15, and 60-view angles, $\lambda = 1$, and $s = 100$, while 6.17(b) shows the corresponding mean activity estimates obtained using MLEM. In each case the mean activity estimate was normalized to the number of view angles. The CP method behaved similarly to the simulations, in that for $\lambda = 1$ and $s = 100$, the activity estimates obtained with 3-view angles were lower than those obtained using 15 and 60-view angles. For MLEM the activity estimates were consistently higher than the estimates obtained using the CP method over the range of activity levels.

The convergence metrics described in chapter 5 were also examined and are shown in figure 6.18 using $s = 100$ and $\lambda = 0.5$ to 7. Here, the metrics are plotted for 800 iterations to better see the convergence behavior of the method, though they were computed out to 2000 iterations. The overall behavior of the CP method was found to be similar to the lung phantom simulations. The rate at which the *PPDG* goes to 0 was faster
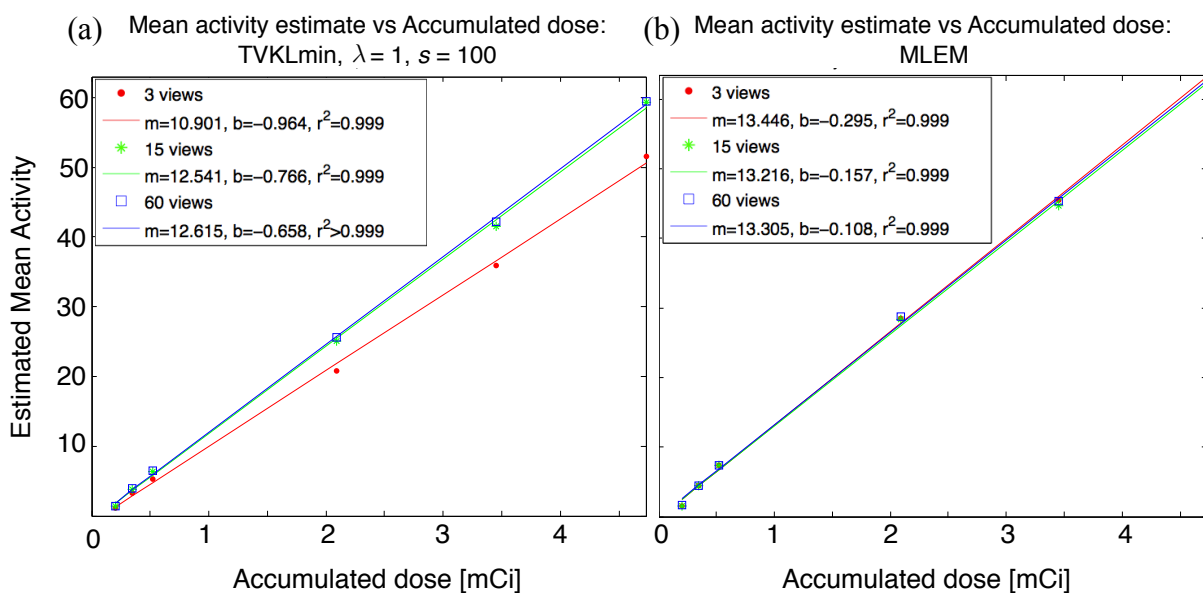
Figure 6.17: Estimated mean intensity in the lungs plotted against accumulated dose for 3, 15, and 60-view angles: (a) CP method using $\lambda = 1$ and $s = 100$, (b) MLEM.
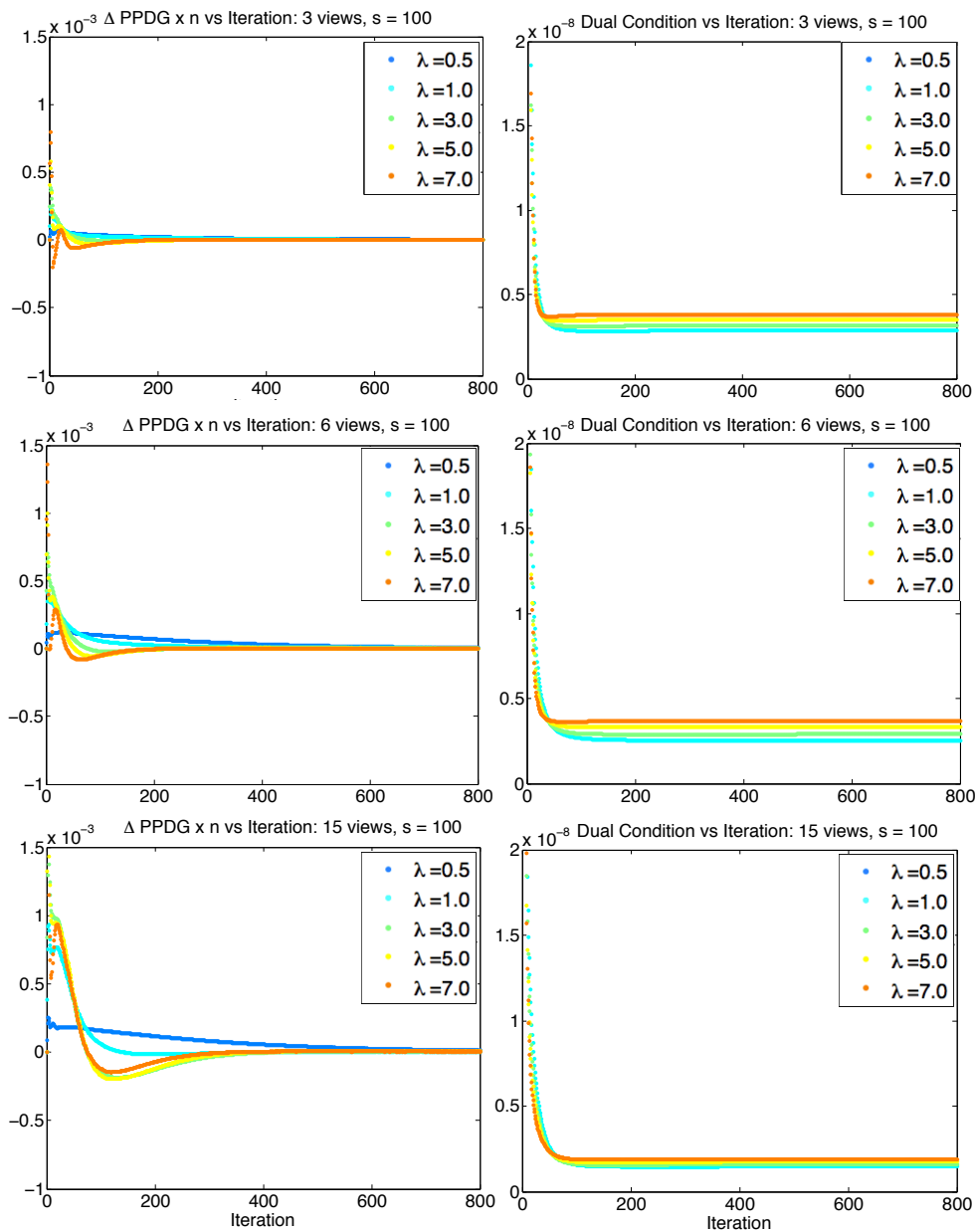
Figure 6.18: The convergence metrics for reconstruction of the 3.45 mCi ^{99m}Tc-MAA injection with $s = 100$. The first column plots $\Delta PPDG \times n$ vs iteration, and the second column plots the dual condition vs iteration. Each row represents 3, 6, or 15-view angles.

than $1/n$ for all λ. However, the *PPDG* does not decrease to 0 at any point between 0 and 800 iterations for $\lambda = 0.5$. The dual condition and the D_{KL} both decay to small values. For all view angles, the data error D_{KL} was smallest for $\lambda = 0.5$. Likewise, as the number of

view angles increased, the dual condition and the data error decreased for all λ up to 800 iterations.

6.4.2 The effect of injected dose

Experimentally, projection data in SPECT is often photon starved due to requirements to reduce dose, reduce the size of the pinhole to increase spatial resolution, reduce the acquisition time for faster scanning, and in some cases, diffuse distribution of the radiopharmaceutical in the organ/tissue. To investigate the effect of changing the injected activity on the CP algorithm, reconstructions were performed on the multi-dose ^{99m}Tc-MAA data set. The CP method was run for 2000 iterations using TV-KL minimization with $\lambda = 1$ and $s = 100$. Figure 6.19 displays the central slice of the reconstructions for the 0.2 mCi, 0.52 mCi, and 3.45 mCi accumulated ^{99m}Tc-MAA dose, where the rows correspond to the number of view angles used to perform the reconstructions. To account for the different intensities in the reconstructions, the window and level were automatically adjusted for each image in figure 6.19 for display purposes.

The CP method was able to reconstruct ^{99m}Tc-MAA distributions that reasonably resembled the lung for all numbers of view angles over the range in activity levels. Steak artifacts were most noticeable for the lowest dose reconstructions, but as the accumulated dose increased the streak artifacts were reduce for all numbers of view angles. At a fixed activity level, a similar trend was observed as the number of view angles was increased.

For comparison, the data were also reconstructed using MLEM with 30 iterations. The results shown in figure 6.20, were consistently noisy compared to their CP counterparts. Although there were little to no streak artifacts observed in the MLEM
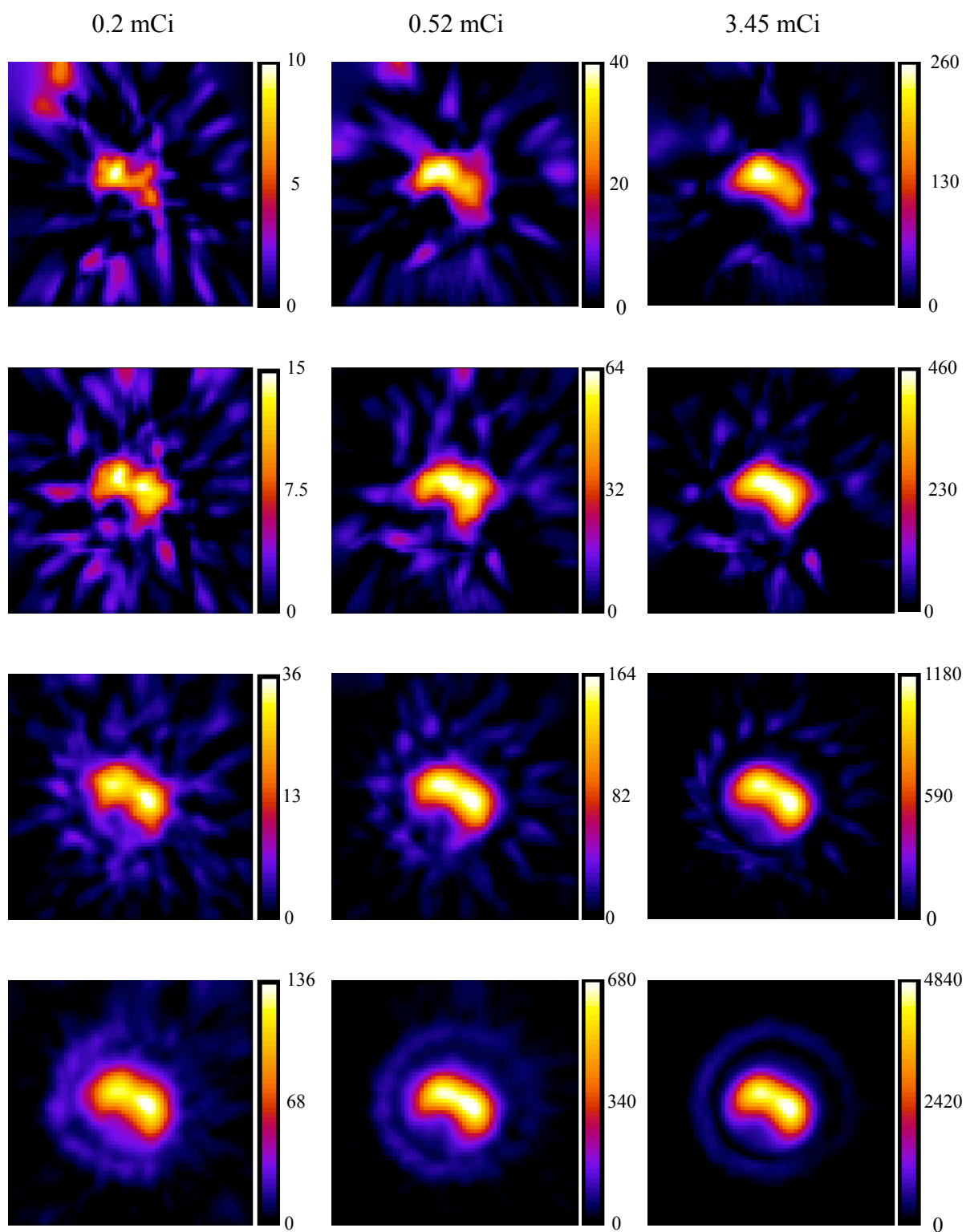
Figure 6.19: ^{99m}Tc-MAA reconstructions using TV-KL minimization: each column represents the accumulated dose, and each row represents 3, 6, 15, or 60-view angles.

0.2 mCi   0.52 mCi   3.45 mCi
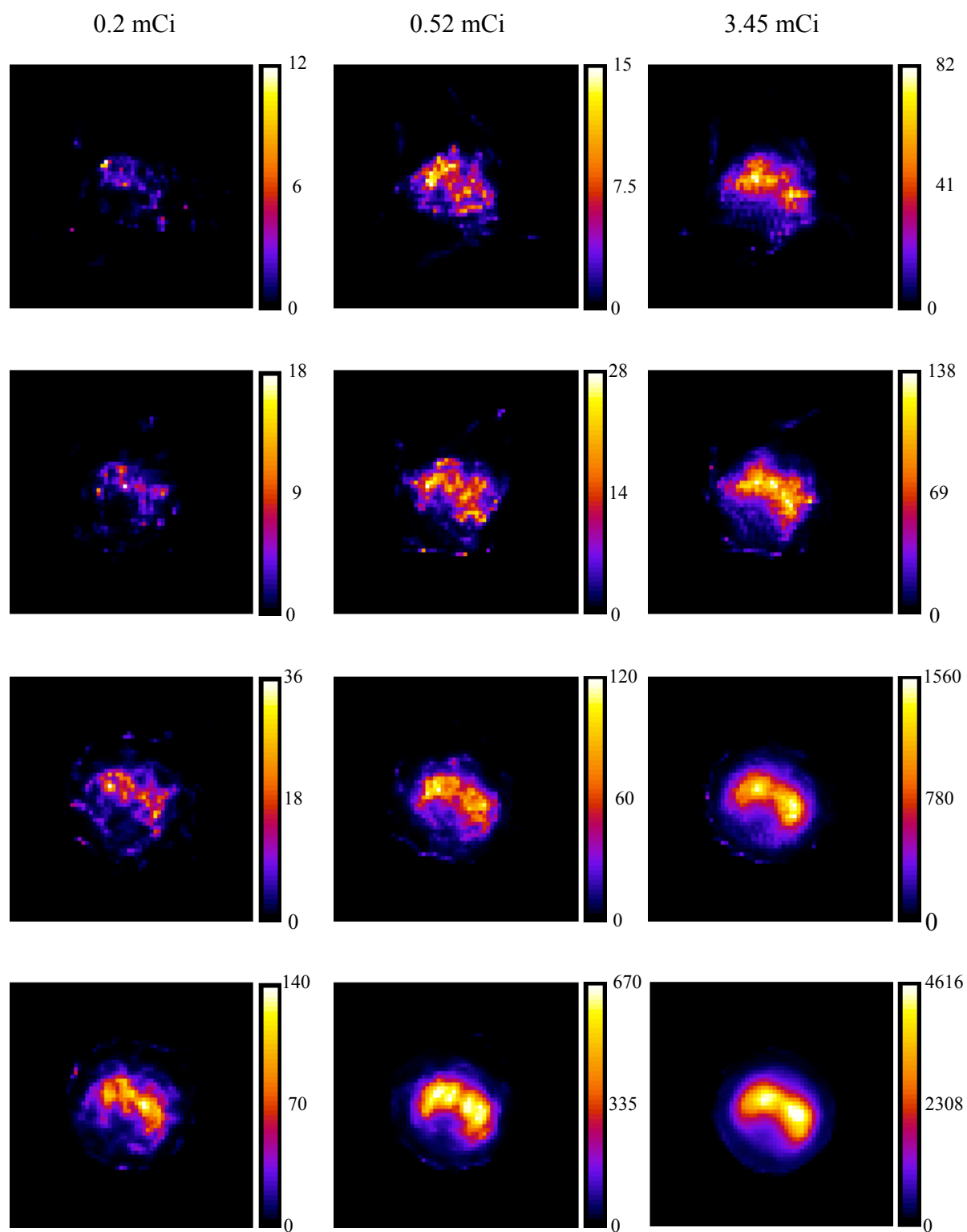
Figure 6.20: ^{99m}Tc-MAA reconstructions using MLEM: each column represents the accumulated dose, and each row represents 3, 6, 15, or 60-view angles.

reconstructions, the method was unable to produce distributions consistent with relatively homogenous ^{99m}Tc-MAA uptake expected in rat lungs for the 0.2 mCi and 0.52 mCi data sets using 3, 6, and 15-view angles. As the amount of activity increased, the lungs were better resolved in the reconstructions. However, even using 60-view angles, the 0.2 mCi reconstruction was relatively noisy compared to the corresponding CP reconstruction in figure 6.19.

The effect of injected dose on the convergence behavior was also investigated. Figure 6.21 shows the convergence metrics with $\lambda = 1$ and $s = 100$ for the entire range of
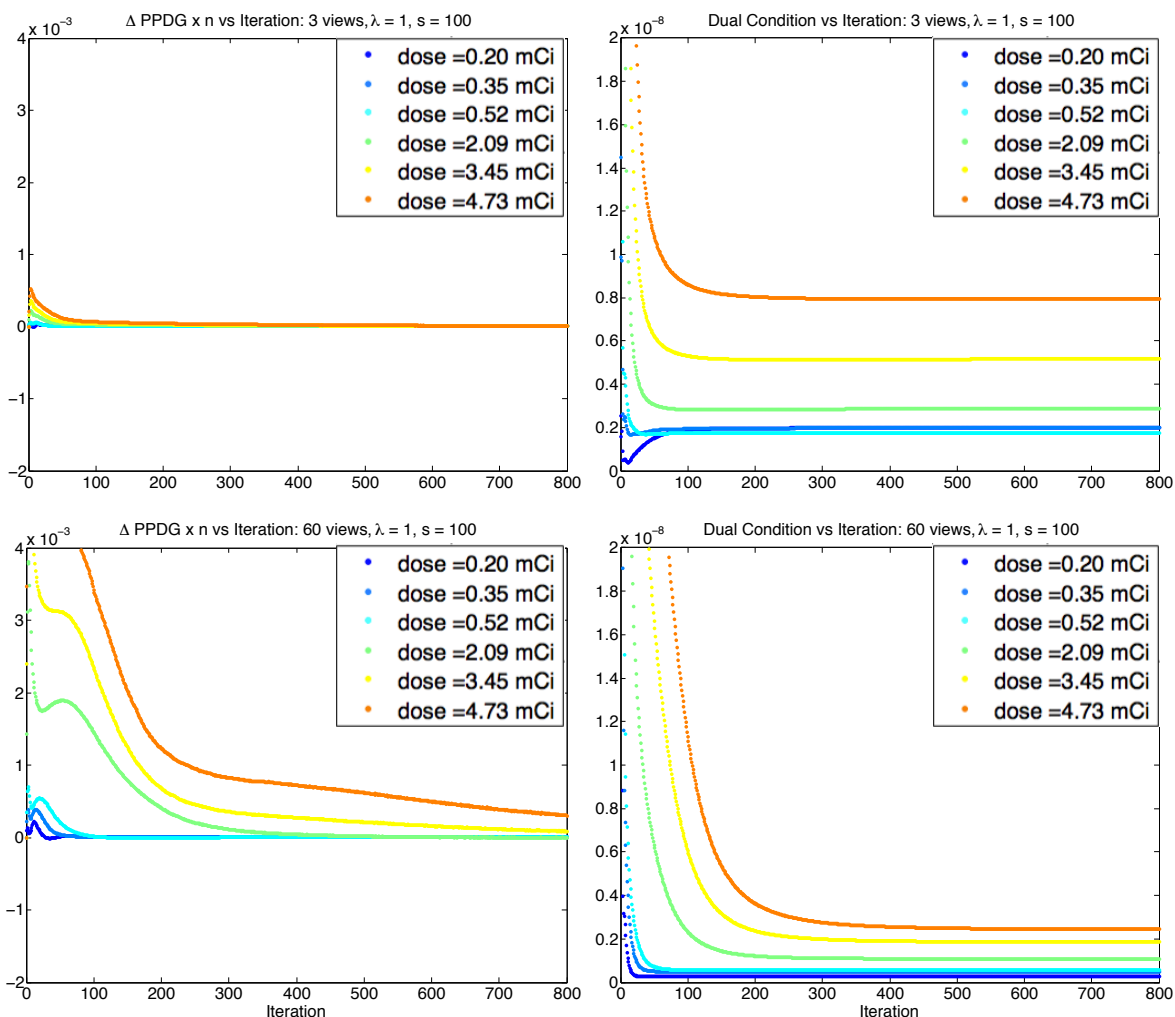
Figure 6.21: The convergence behavior with injected dose. The left column plots $\Delta PPDG \times n$ over iteration, and the right column plots the dual condition over iteration. Each row represents 3 and 60-view angles.

MAA activities. The $\Delta PPDG \times n$ goes to 0 faster for the lowest accumulated dose for all view angles. Likewise, the dual condition is lower after 800 iterations for the lowest accumulated dose for 60-view angles. In the 3-view angle reconstructions, the dual condition was smallest for the 0.35 mCi MAA injection. For 0.2 and 0.35 mCi injections, a local minima was observed in the dual condition before leveling out. Overall, Figure 6.21 demonstrates that the convergence metrics were sensitive to the amount of activity being reconstructed, although the general trend of the metric was similar in all cases.

6.4.3 The effect of changing the scaling parameter *s* on activity estimates

We also investigated the effect of increasing s on the *in vivo* MAA reconstructions. As in the case of chapter 5, TV-KL minimization was performed using the CP algorithm run for 2000 iterations. The results are shown in figure 6.22 for $\lambda = 1$ and $s = 1$, 100, and 500. Increasing the scaling parameter had the effect of reducing the streak artifacts in the reconstructions, as was the case in the simulations. However, for $s = 500$, there was no noticeable reduction in streak artifacts.

Likewise, the effect of s on the mean activity estimates are plotted against the accumulated dose, figure 6.23, for 3 and 15-view angles and $s = 1$, 100, and 500. The mean intensity estimates were lower for 3 and 15-view angles for the $s = 1$ case. Larger estimates were obtained for $s = 100$ and $s = 500$. There was no apparent difference between the $s = 100$ and $s = 500$ mean activity estimates, consistent with the observations in figure 6.22.
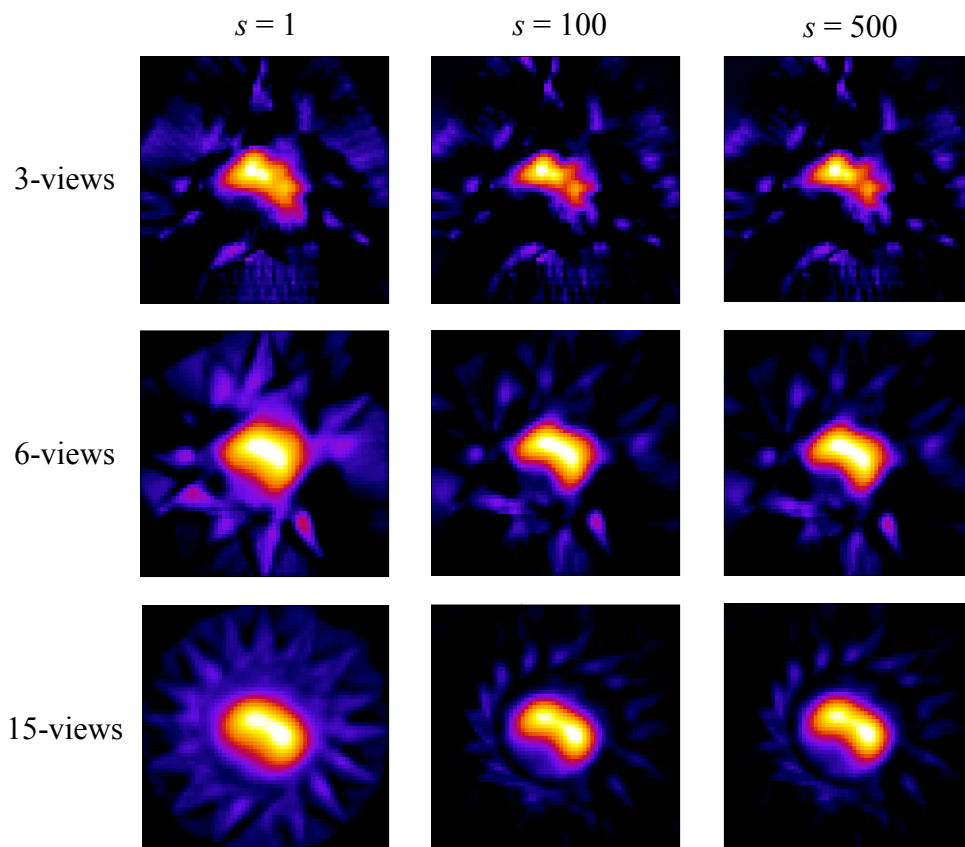
Figure 6.22: Central slice reconstructions using $\lambda = 1$. The columns represent $s = 100$, 200, and 400, and the rows represent 3, 6, and 15-view angles.
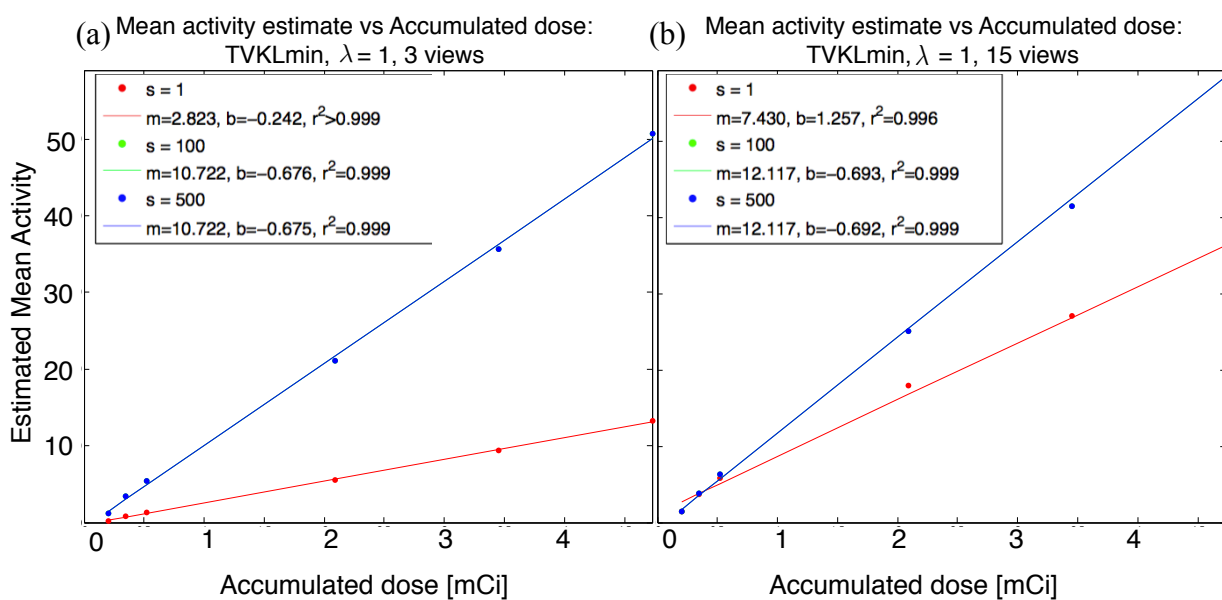
Figure 6.23: Estimated mean intensity in the lungs plotted against accumulated dose for $s = 1$, 100, and 500: (a) 3-view angles, and (b) 15-view angles.

6.5 Hyperoxic lung study

Reconstructions were performed on the HMPAO data, obtained using the protocol described in 6.2.3, using the CP method and MLEM. MAA data obtained from the same animal was reconstructed using 60-view angles, the lung boundaries were identified using the segmentation procedure outline in 6.4.1, then the mean activity within the lung boundaries of the HMPAO reconstructions was calculated. The uptake of HMPAO within the rat is not limited to the lung region, as is the case for MAA. Hence, the task here was to reconstruct activity distributions with a relatively low signal compared to background. In fact, previous studies have shown that only about 3-13% of the HMPAO injected dose accumulates in the lungs [6].

6.5.1 HMPAO reconstructions

Using observations gleaned from the MAA results, the HMPAO data was reconstructed using $\lambda = 1$ or $\lambda = 3$, $s = 100$, and for 3 to 60-view angles. Additionally, to eliminate the remaining streak artifacts, the system matrix was calculated by masking the reconstruction volume to a cylindrical region whose diameter was slightly greater than the diameter of the animal holder. As a consequence of masking the system matrix, the back-projector's response at the boundary of the mask is treated like an edge feature in the reconstruction, which results in the halo artifact around the reconstruction. After reconstructing the data, the reconstruction volume was masked with a cylindrical volume with a diameter equivalent to the animal holder. This process is illustrated in figure 6.24.

Examples of 3 and 60-view HMPAO reconstructions are shown in figure 6.25, where the activity is not just localized to the lungs. The choice of λ was found to largely
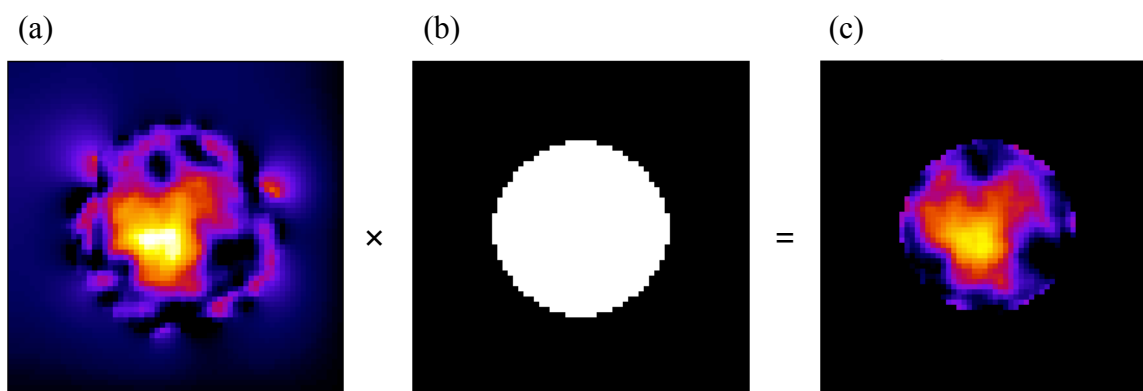
Figure 6.24: HMPAO masking procedure: (a) reconstruction incorporating a cylindrical mask into the system matrix, (b) the animal holder mask, (c) the post-processed volume.
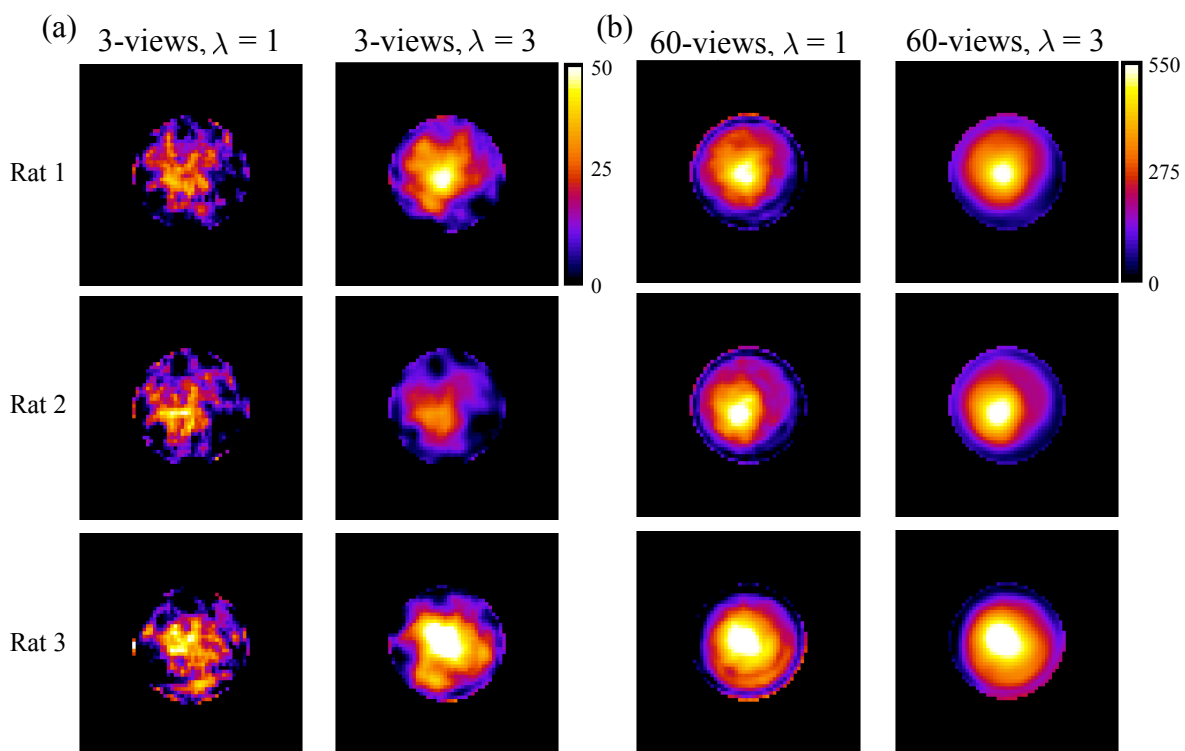
Figure 6.25: HMPAO reconstructions: (a) 3-view reconstructions, (b) corresponding 60-view reconstructions. Each row represents a different dataset, with $s = 100$.

depend on the number of view angles. For the 3-view reconstructions, the $\lambda = 1$ reconstructions were extremely noisy compared to the $\lambda = 3$ case (the first column in 6.26(a)). Conversely, the 60-view reconstructions were blurry for $\lambda = 3$.
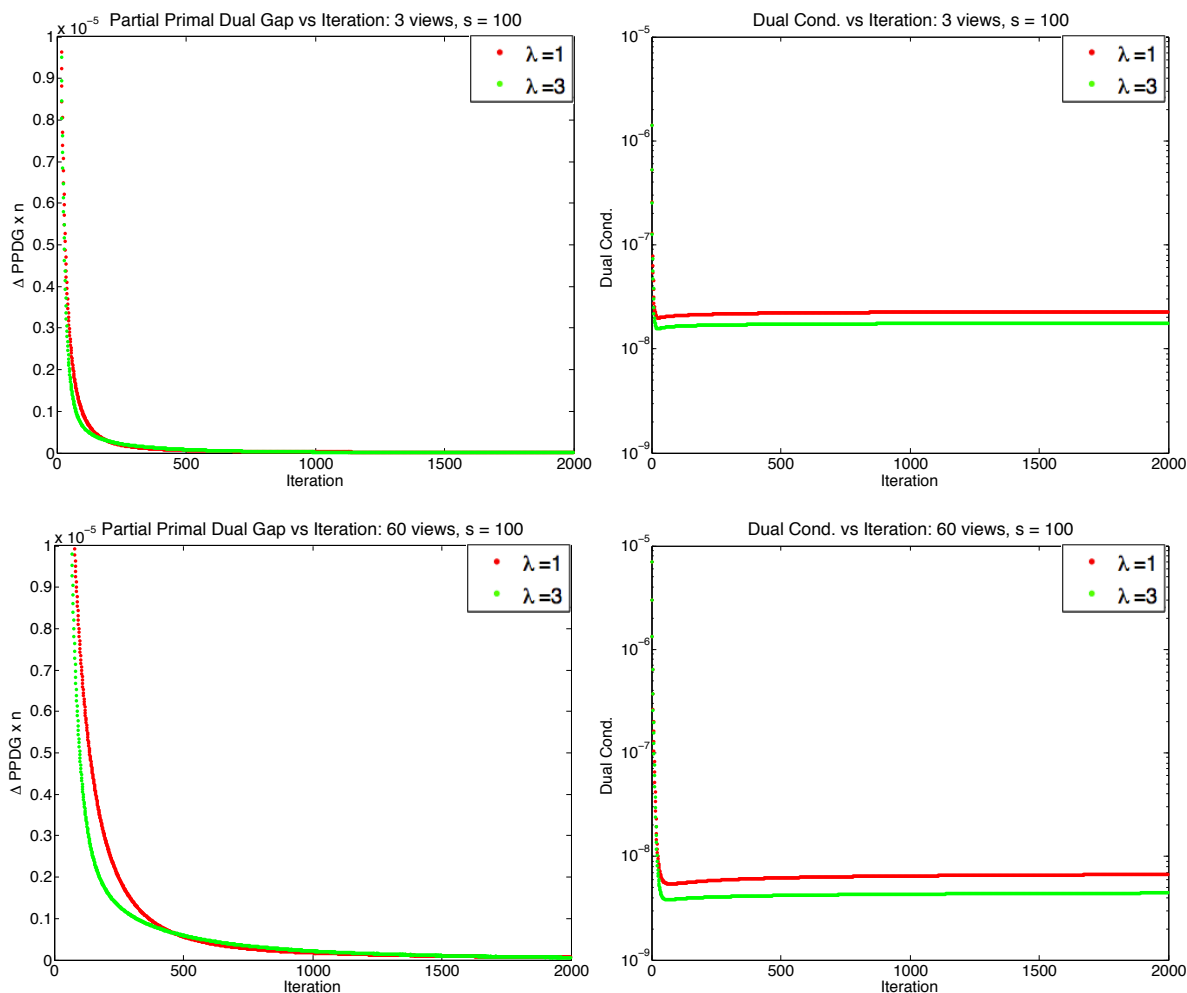
Figure 6.26: HMPAO reconstruction convergence metrics. The first column plot $\Delta PPDG$ × n over iteration, and the second column plots the dual condition over iteration for $s = 100$.

Figure 6.26 shows $\Delta PPDG \times n$ where the rate of change of $PPDG$ decayed to zero faster than $1/n$ for both 3 and 60-view angles and for both $\lambda = 1$ and $\lambda = 3$. The dual condition also decayed to a small value, with a slight minimum observed before leveling off, consistent with the low dose MAA injections in figure 6.21. In this example, the CP method converged faster for 3-view angles than 60-view angles.

In order to quantify HMPAO uptake in the lung regions, the segmentation

procedure in 6.2.3 was applied to the MAA reconstruction to create a preliminary lung

mask as shown in figure 6.27. However, because of the proximity of the heart and liver to
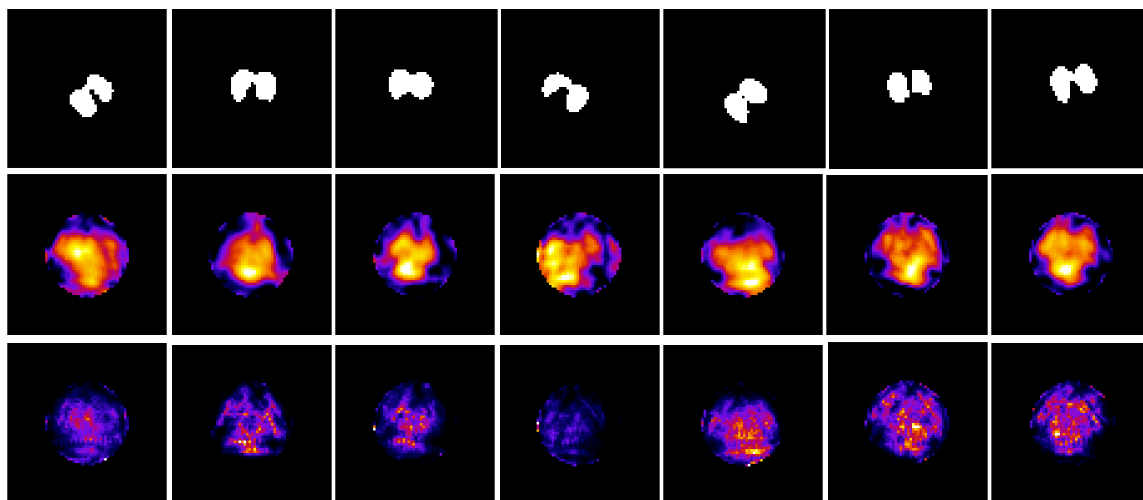
Figure 6.27: Row 1 shows the lung masks segmented from the ^{99m}Tc-MAA acquisition, row 2 shows the corresponding CP reconstructions, and row 3 shows the corresponding MLEM reconstructions.

the lungs, it was hypothesized that the lung region contained activity from these other

organs in the HMPAO reconstructions as a consequence of the 2mm pinholes. This was

investigated by performing an empirical threshold on the HMPAO reconstructions, and

examining the regions of high activity, as described in Appendix D. In figure 6.28, a

mask of the high activity, in blue, was created and overlaid onto the lung masks, in red. In

general, the regions of high activity were consistent with the location of the heart and

liver. 3D renderings of two examples of the composite lung/heart masks are shown in
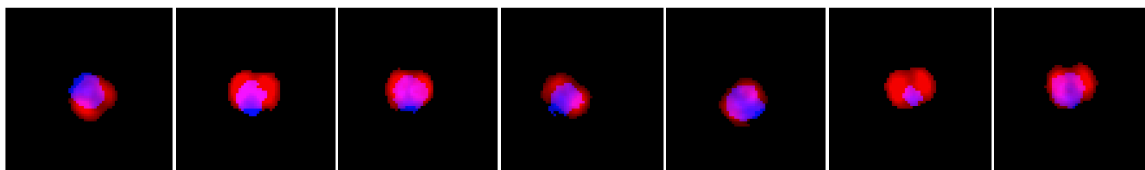
Figure 6.28: Overlay of the lung mask, in red, and the heart mask in blue.

figure 6.29. The lung mask was further refined by excluding regions that overlap with the heart and liver, to correct for liver contribution that was visible in the composite volumes (the left image in figure 6.29).
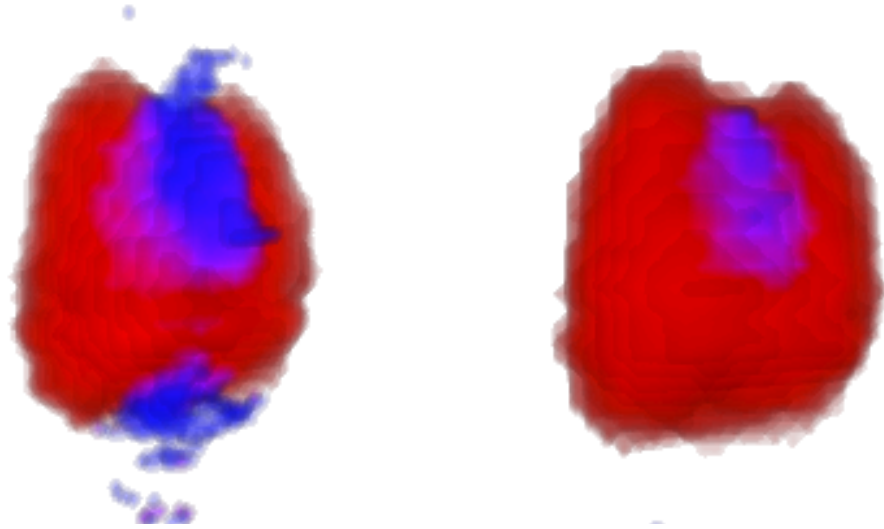
Figure 6.29: 3D renderings of the lung and heart composite masks from figure 6.28.

The final lung mask was applied to the HMPAO reconstructions, and the mean intensity in the lung region was calculated. The mean intensity in each lung region was normalized to the number of view angles, the acquisition time, the number of nonzero voxels in the lung mask, and the injected dose to obtain the final mean activity estimate in the lung region. Figure 6.30 shows group averages obtained from 4 control and 4 hyperoxia-exposed rats using the CP method and MLEM. For all view angles, $\lambda = 3$ and $s = 100$.

Since the uptake of HMPAO is hypothesized to be a biomarker for oxidative stress, in quantifying the difference between the control and 24 hr hyperoxic groups the task is to determine if early lung injury can be detected using our reconstruction methods.
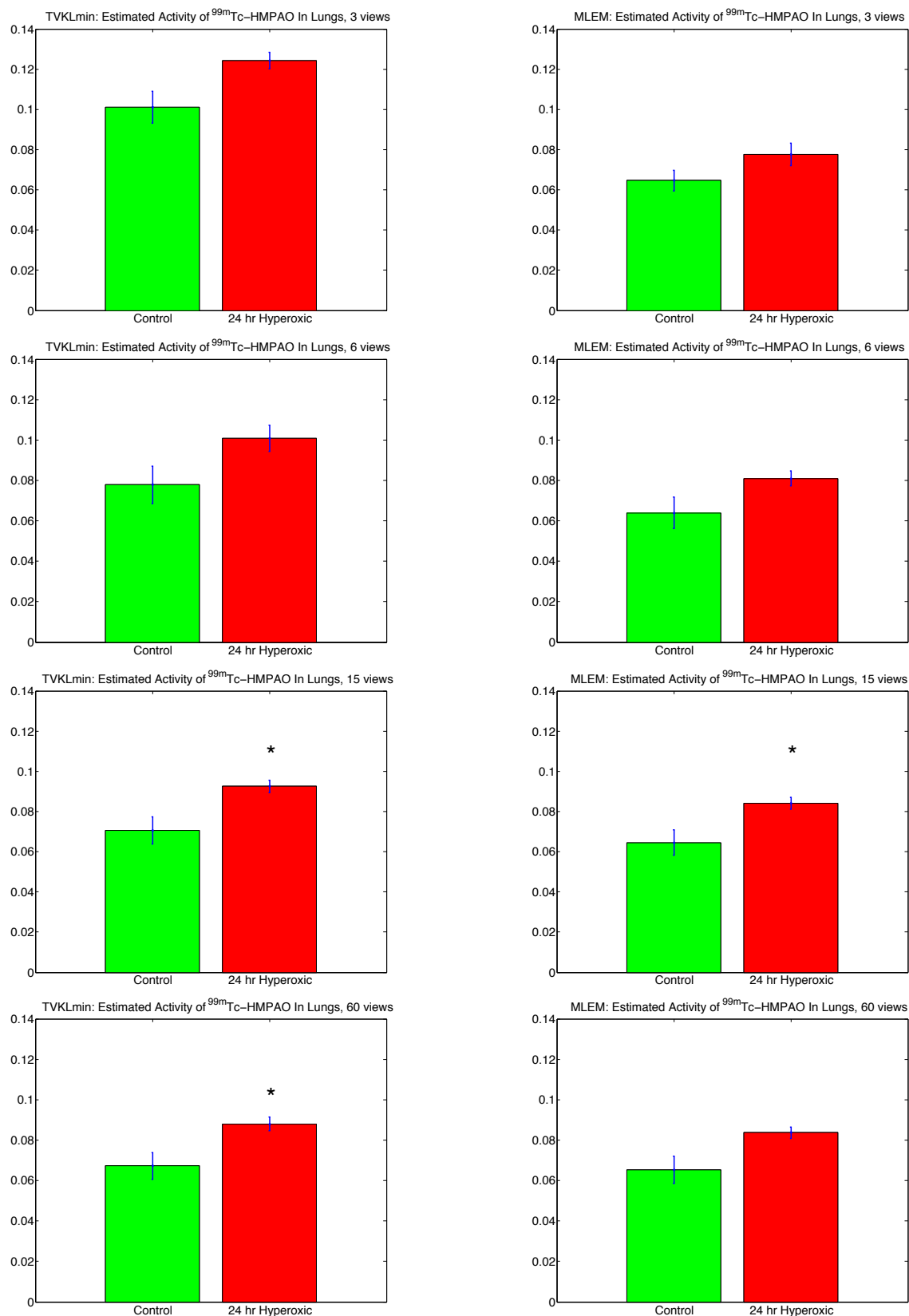
Figure 6.30: HMPAO mean activity estimates in the lung regions for control (green) and hyperoxic (red) rats. The left column was obtained using the CP method, and the means on the right were obtained using MLEM. $n = 4$ in all cases. * indicates $p < 0.05$.

Overall, estimates of the mean activity of in the lungs of the hyerpoxic group were ~20%
higher than the control group. This difference was tested using an unpaired *t*-test [66] on
the means, and statistically significant differences were found in the case of 15 ($p =$
0.041) and 60-view angles ($p = 0.043$) using the CP method and 15-view angles ($p =$
0.0426) using MLEM.

6.6 Discussion

This chapter demonstrated the use of the CP method with a TV-penalty on
experimental SPECT data acquired in our laboratory. Extending the reconstruction
methods described in the previous chapter required additional preprocessing steps to
reduce artifacts in the reconstructions. Discrepancies between the forward model and the
experimental projections due to misalignments, background activity, and center-of-
rotation offsets were accounted for. Additionally, a normalization routine was developed
that mitigates ring artifacts due to the back-projector's response to a flood projection and
models the intensity response of each pinhole projection individually.

The results from the multi-dose ^{99m}Tc-MAA data suggest the CP method works
well over a large range of injected activity levels and can be extended to experimental
few-view data. Mean activity estimates in the lung regions increased linearly with
accumulated injected dose for both reconstruction methods. The slopes in figure 6.15
using the CP reconstructions were smaller compared to MLEM, unlike the simulations.
This difference may be due to unaccounted for experimental errors such as detector
yellowing [79], a slight wobble in the axis of rotation, or additional misalignments.
However, the convergence behavior is consistent with the results from chapter 5, where

alignment errors were not present. Additionally, with an appropriate choice of λ and s, the CP method is capable of reconstructing activity distributions consistent with rat lung anatomy, even for the smallest accumulated MAA dose used in section 6.4 (0.15 mCi).

The CP method was also able to reconstruct ^{99m}Tc-HMPAO data for which the activity level in the lungs was low and nearly the same as the surrounding organs and tissues. In figure 6.25, larger smoothing parameters were required to control the amount of noise introduced to the reconstructions using 3-view HMPAO data. Conversely, smaller smoothing parameters were required to prevent blurring in the 60-view reconstructions. Using $\lambda = 3$, the 3-view activity distributions were similar to the 60-view reconstructions using $\lambda = 1$. In figure 6.27, the 3-view reconstructions of HMPAO using MLEM were noisy compared to the CP reconstructions.

Finally, the methods were applied to a lung study comparing the uptake of HMPAO in control and hyperoxic rats. Activity in the 24 hr hyperoxic lungs was approximately 20% higher than the controls on average. This difference is qualitatively consistent with previous results obtained using planar HMPAO imaging [5, 24]. A direct comparison cannot be made, however, since in the planar imaging study the ratio of the activity in the lungs to the activity in the blood was used to compare the two groups. The "blood" activity was obtained by finding the mean activity in the forearm of the animal, which was outside the field-of-view in our study. Another potentially confounding factor is that in our study the rats were positioned vertically, compared to the horizontal positioning and parallel hole collimator used in the planar study. Changes in orientation result in changes in regional blood flow and ventilation, due to hydrostatic pressure [61], which may affect the lung volume [35] and hence HMPAO uptake in the lung. One

potential means to investigate the impact of these zone conditions would be to image each rat in both the horizontal and vertical position. Alternatively, using the existing setup, drugs that increase the heart rate could be used to increase blood flow in the lung might result in more homogeneous pulmonary perfusion.

Although statistical differences between the groups could not be established for 3 ($p = 0.050$) or 6-view angles ($p = 0.094$) using the CP method, the approach to quantification presented here can be refined to reduce the variation in the mean estimates. Experimentally, the error on the mean may be due to slight variations in the camera response between acquisitions. Figure 6.31 shows a plot of the estimated mean intensity
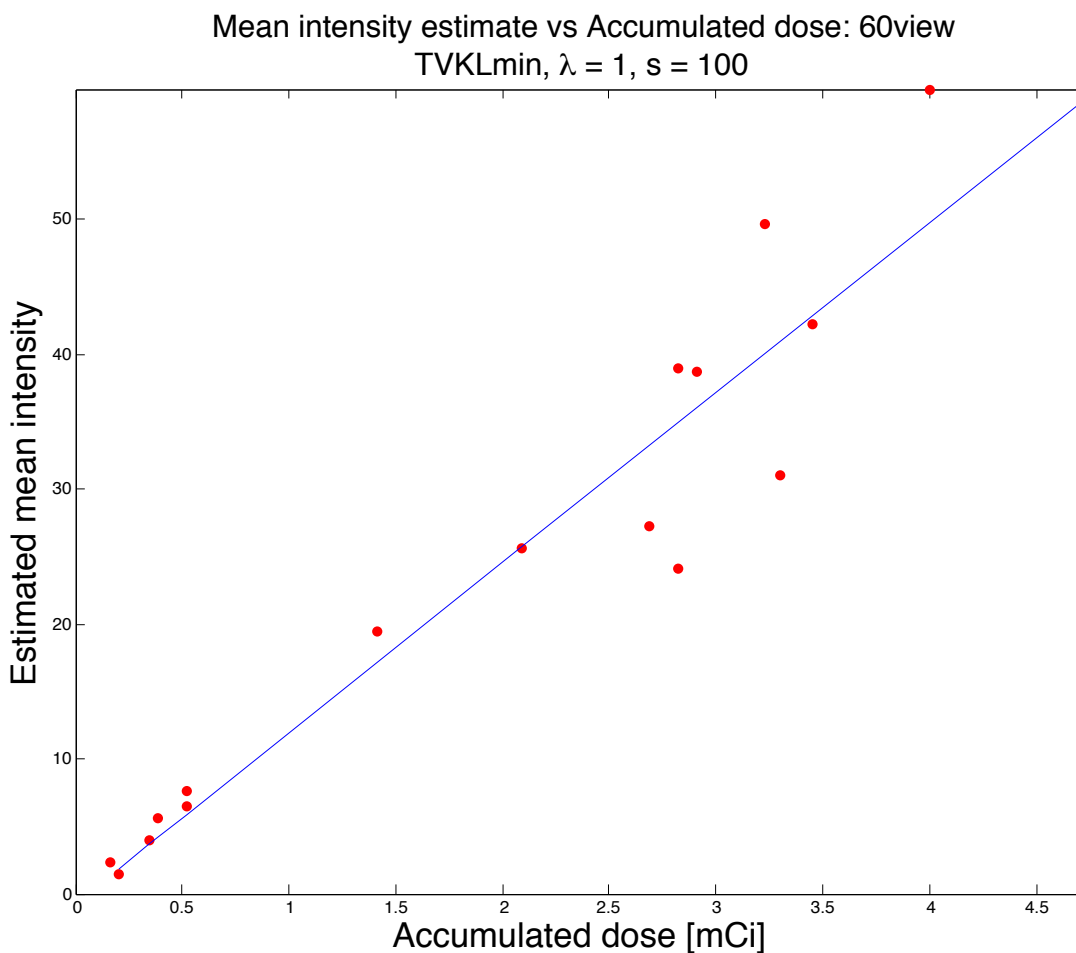
Figure 6.31: Mean intensity estimates in the lungs for all ^{99m}Tc-MAA data acquired from sections 6.4 and 6.5.

plotted against accumulated dose for all the MAA rat data acquired from sections 6.4 and 6.5. The best fit line from figure 6.17(a) is plotted in blue for comparison, and there is a variation in the camera's response to the injected activity. However, at accumulated doses observed in lung HMPAO imaging (≤ 0.5 mCi), the variation in the camera's response appears much smaller. In the future, the imaging protocol should be adapted to account for the camera's response by imaging a multi-dose calibration phantom with each rat experiment.

Chapter 7

Conclusion And Discussion

This dissertation describes an iterative approach for SPECT reconstruction using the Chambolle-Pock algorithm. To reduce the size of the system matrix, well-matched rotation-based forward- and back-projectors were designed to take advantage of fast matrix-vector multiplication routines implemented on GPUs. When applied to TV-penalized D_{KL} minimization, the method's convergence behavior was characterized on fully 3D simulations using few-view data. Furthermore, the methods were extended to *in vivo* rat lung imaging studies. This SPECT reconstruction approach grew out of work done in x-ray CT by Sidky, et al [91, 92, 93]. Although other approaches to few-view SPECT have been developed [48, 104], to the best of our knowledge, this is the first application of the CP algorithm to fully 3D SPECT using a TV penalty. This chapter summarizes the major highlights from this document, provides suggestions for future work, and concludes this dissertation.

7.1 Major Highlights

7.1.1 Implementation of matrix-based reconstruction methods

Chapters 2 and 3 present our approach to the SPECT reconstruction problem. SPECT reconstructions were performed by pre-calculating a system matrix that models the imaging system's collimator and imaging geometry. In this way the forward- and back-projection operations were implemented to take advantage of fast and established matrix multiplication routines. To this end, a ray-driven model was used to calculate the

system matrix using a pinhole mask that models the dimensions of the pinholes, as proposed by Wietholt [102], and was extended to multi-pinhole collimators. The system matrix is calculated for a single view angle as a means to reduce the size of the system matrix. As such, a complete forward-projection operation may be performed by digitally rotating the reconstruction volume and multiplying the volume by the system matrix. The size of the system matrix was further reduced by only storing the nonzero components.

Chapter 4 describes emission CT and the methods used to analyze our matrix-based implementation. The problem was cast as an optimization problem and two iterative solution methods were presented. An established approach, MLEM, generates a series of successively better approximations to the reconstruction problem by minimizing D_{KL} data fidelity. Since MLEM relies solely on minimizing data fidelity, it can be sensitive to noise. Conventionally, MLEM is stopped after 30 iterations to reduce the amount of noise introduced to the reconstruction, potentially limiting how well the method performs with few-view data. Thus we investigated the Chambolle-Pock (CP) method that incorporates a smoothness constraint into the minimization procedure, thereby penalizing noisy solutions. The CP method is a convex optimization routine, so that convex noise penalties can be incorporated. This is advantageous since penalties that promote sparsity in some basis of the solution can be used with a compressed sensing strategy to solve very undersampled linear systems [15, 99], suggesting its potential use in few-view SPECT.

In chapter 5, MLEM and the CP method were implemented using our matrix-based approach. A simple 3D simulation was used to validate that the forward- and back-projection operators were well matched in the CP method using an L_2 fidelity metric and

a quadratic penalty term. The solutions converged rapidly to the exact solution using 15 view angles, suggesting our matrix based implementation is well suited for the CP method.

The method was investigated further using a more realistic lung phantom that was created based on an experimental micro-CT scan of a rat thorax with heterogenous uptake in the lung regions and the surrounding tissue. Pseudo-data was generated by applying the forward-projector to the digital lung phantom and incorporating Poisson noise. The data were reconstructed using the Kullback-Leibler data error (D_{KL}) metric with a TV (total variation) smoothness penalty. The solutions did not converge faster than $1/n$, but the lung phantom was reconstructed accurately using an appropriate choice of the smoothing parameter and primal-dual step size. Furthermore, the mean intensity within a lung ROI was recovered accurately. Table 7.1 summarizes the impact of each control parameter for TV-KL minimization, and suggests a reasonable range of smoothing parameter λ and scaling step size s if only a small number of view angles are used to perform the reconstruction.

Simulation parameters	Impact on reconstruction	CP optimal value
N_θ (number of view angles)	Controls image fidelity	$N_\theta \geq 3$
λ (smoothing parameter)	Controls image smoothness	$0.01 \leq \lambda \leq 1$ (no noise) $0.1 \leq \lambda \leq 1$ (5 % noise)
s (step size parameter)	Controls convergence rate	$50 \leq s \leq 200$

Table 7.1: Simulations: impact of each control parameter on the phantom reconstructions.

7.1.2 Application to *in vivo* rat lung studies

To evaluate the utility of the CP method (TV-KL minimization) for experimental animal data, we performed a pilot rat imaging study. Chapter 6 discusses techniques used to tune the theoretical forward-projection model to better match the experimental setup by accounting for misalignments, background noise, and the back-projector's flood response. The resulting matrix-based CP implementation was robust enough to reconstruct the pulmonary perfusion marker ^{99m}Tc-MAA over a large range in activity accumulation in the lungs, consistent with simulation results from Chapter 5. Likewise, the convergence behavior of the method was consistent with that observed in the simulations. The CP method was able to reconstruct activity distributions consistent with rat lung anatomy even with only 3 view angles. MLEM reconstructions, in comparison, were noisy when fewer than 60 view angles were used.

In a second rat study, we investigated the use of ^{99m}Tc-HMPAO as a biomarker of oxidative stress that occurs in acute lung injury. Rats were exposed to either 95% O_2 (hyperoxia) or room air (control) for 24 hrs. We then administered ^{99m}Tc-HMPAO to the anesthetized rats, imaged, and 3D reconstructions were performed on the data using our CP implementation. The boundary of the lungs was obtained from a ^{99m}Tc-MAA reconstruction from the same animal, and the mean activity within the lung region, excluding the heart and liver, was calculated. Results confirmed that ^{99m}Tc-HMPAO uptake in the lungs, obtained using 15 or more projection view angles, was higher in the hyperoxic group than the control group. These results were obtained using 4 animals per group so that increasing the number of animals may also lead to a significant difference when a smaller number of views are used.

Table 7.2 summarizes the task oriented results/recommendations for estimating

the mean activity within an ROI of the reconstructions using the CP method and MLEM.

Overall, the mean activity estimated from the CP reconstructions was insensitive to the

choice of smoothing parameter λ, and the method converged with fewer iterations when

the scaling between the primal and dual step sizes was set at $s = 100$. Although it takes

the CP method 0.018 s per iteration, compared to 0.010 s per iteration using MLEM, the

CP method was better at reconstructing lung-like distributions using only 3 and 6 view

angles and performed better on very low dose projection images. In the experimental

Imaging Task: Lung Mean Activity Estimation					
	N_θ	λ (smoothness)	s (step size scaling)	N_{iter}	Dose
CP:	≥ 6	Insenstive	100	100-200 (Total runtime = 1.8-3.7s)	≥ 0.5 mCi
MLEM:	≥ 15	-	-	30 (Total runtime = 0.3s)	≥ 3 mCi

Table 7.2: Optimal imaging parameters for estimates of lung activity in ROI. N_θ is the number of view angles and N_{iter} is the number of iterations.

reconstructions, convergence of $\Delta PPDG$ was observed as a theoretical measure of

convergence. However, it is reasonable to use the dual condition and D_{KL} as practical

measures of convergence. In that case, the experimental results in figures 6.18 and 6.26

suggest that 100-200 iterations is sufficient to obtain reconstructions that do not change

appreciably with more iterations.

7.2 Future Work

A ray-driven method was chosen for this work because it can easily be adapted to

account for many physical aspects of the imaging system and it is computationally

efficient, making it suitable for iterative reconstruction methods. Thus, future work could include incorporation of additional physical aspects of the SPECT imaging system, e.g. re-weighting the system matrix **H** may be used as a means to account for pinhole edge penetration [95, 102, 107] or for photon attenuation within the imaged activity distribution [25]. If the boundaries of the activity region within the field of view were known a priori, either from the ^{99m}Tc-MAA images or from x-ray CT images, those could also be incorporated as a prior into the total-variation when using TV-KL minimization in the SPECT reconstruction problem.

Another avenue for future work is the investigation of the use of different noise penalties [91]. Although TV penalization was demonstrated *in vivo*, a penalty that is not as sensitive to transitions in coverage in the back-projector's response could potentially reduce streak artifacts present in the reconstructions. It was found that using an L_2 noise model to reconstruct Poisson noisy data results in reconstructions which are not sensitive to the smoothing parameter. However, in conjunction with different smoothness constraints, fidelity measures that differ from the noise model might also be considered. Additionally, different noise models might also be considered. For example, a binomial noise model might be considered for low dose acquisitions.

For our simulations and experimental study, a 5-pinhole collimator with fixed image-to-pinhole and pinhole-to-center of rotation distances was used. It would be interesting to adjust the imaging geometry to increase the number of voxels in the reconstruction volume that are visible from all view angles and all pinholes. It is expected that this would reduce the number of transitions in the back-projector's response, leading to reduced ring artifacts. The drawback to adjusting the imaging distances is that the

object volume is further from the detector, which could result in more scattered photons

and a reduction in count rate. More activity could be used to compensate if the goal is to

perform rapid acquisition. Alternatively, different collimators could be constructed for the

current system geometry. The current 5-pinhole collimators were constructed to image

the lung region of a rat. When the activity distribution extends beyond the lungs, as is the

case with the ^{99m}Tc-HMPAO data, a collimator with full coverage of the animal holder

would improve the back-projector's response in the regions of interest in the

reconstruction volume. In either case, the TV-penalty is less likely to treat transitions in

the back-projector's response as edge features. To this end, a possible design

improvement would be to put the line-of-sight of the off-center pinholes at an angle, so

that the line-of-sight for all pinholes is centered on the center-of-rotation. The existing

pinholes might also be modified to stretch the large bore diameter in the x-direction in

one quadrant of the pinhole. This would increase the acceptance angle of the off-center

pinholes to accommodate a larger field-of-view, as illustrated in figure 7.1(a) and (b).
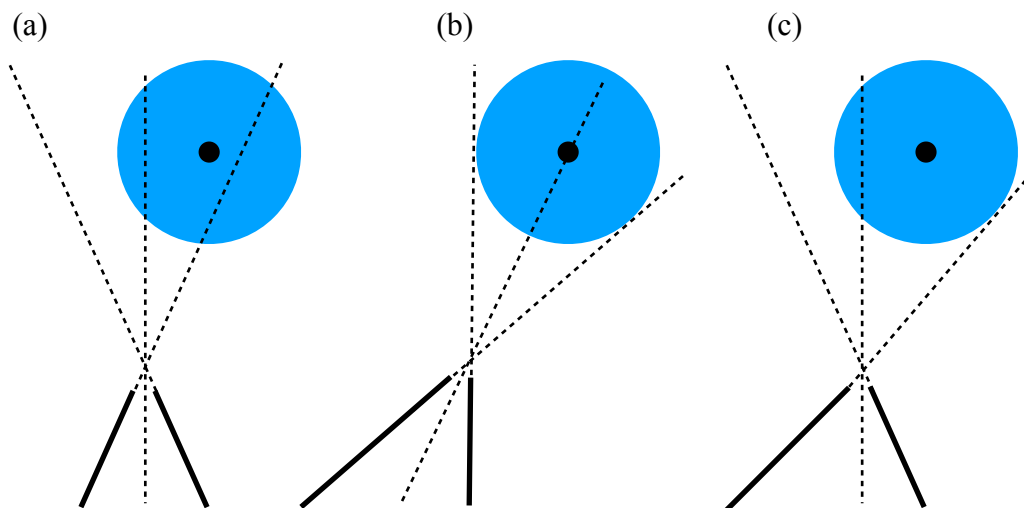
Figure 7.1: (a) Off axis pinhole, (b) off axis pinhole bored at an angle, and (c) modified
pinhole from (a). The blue circle represents the animal holder, the dotted lines
represent the line of sight, and the solid lines represent the pinhole cross-section.

One ongoing challenge to the current implementation of the CP method is the reduction of streak artifacts outside the regions of activity when only a few view angles are used. Although improvements in the alignment of the system may reduce these artifacts, future implementations may take advantage of the CP method's linear update to reduce the activity assigned to regions outside of the region of activity. Improved differentiation between the background and the activity regions might be achieved by introducing a multiplicative factor to the back-projector. Numerically, the activity distributions would have more energy than the projections. However, the method would assign more weight to voxels with a true signal. A naive multiplicative factor can be applied to the back-projector by modifying equation 6.1 to

$$h_{ij} = h_{ij} + \left(1 - w \cdot (\mathbf{H}^T \mathbf{1})_j\right).$$

(7.1)

This is demonstrated in figure 7.4. where ^{99m}Tc-MAA reconstructions using the original **H** matrix from Chapter 6 and those using the H from equation 7.1 are compared. This preliminary result indicates that additional modifications to the system matrix and the normalization protocol would be required to avoid introducing additional ring artifacts. This example illustrates that future modifications can be easily incorporated into the system matrix to better suit the system design or TV-penalty.

Experiments involving dynamic SPECT imaging must employ multiple detectors simultaneously. In order to fully utilize the system matrix strategy developed in this dissertation there must be a well-defined system geometry in order to use the same system matrix for all gamma cameras in the system. The gamma cameras must be aligned well with their respective collimators and to the system's center-of-rotation. Likewise, the

(a)   (b)
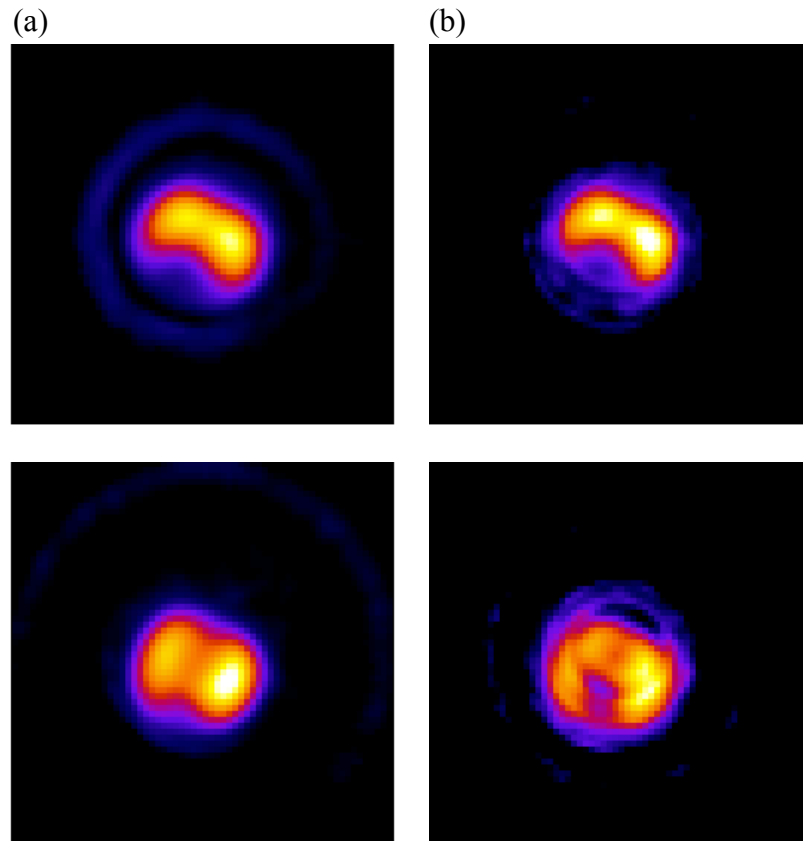
Figure 7.2: Two CP cross-sections of the same MAA data: column (a) uses the matrix
from chapters 5 and 6, column (b) uses a system matrix with multiplicative factor
built into the back-projector.

distance between the image plane and the distance between the collimator and the center

of rotation must also be identical for all cameras. In general, small deviations in the

alignment of the SPECT system may be accounted for by performing an additional

calibration to the procedure described in Appendix A. This procedure involves physically

measuring the contribution of each voxel to a projection by placing a gamma-ray point

source at every location in the physical reconstruction space and imaging the activity

through a collimator [21, 34, 43]. This approach would require as many projections in the

system matrix as cameras in the system, demanding more memory than is currently

available on a single GPU. However, given the proper alignment, a system with multiple cameras may utilize the methods developed here. Many GPUs may then be used to simultaneously reconstruct multiple dynamic volumes. Additionally, as the memory and thread size of GPUs increases, the current implementation may be readily extended to larger reconstruction volumes. Sparse block matrix-vector multiplication routines may also be used to improve performance as the size of the matrix increases [74].
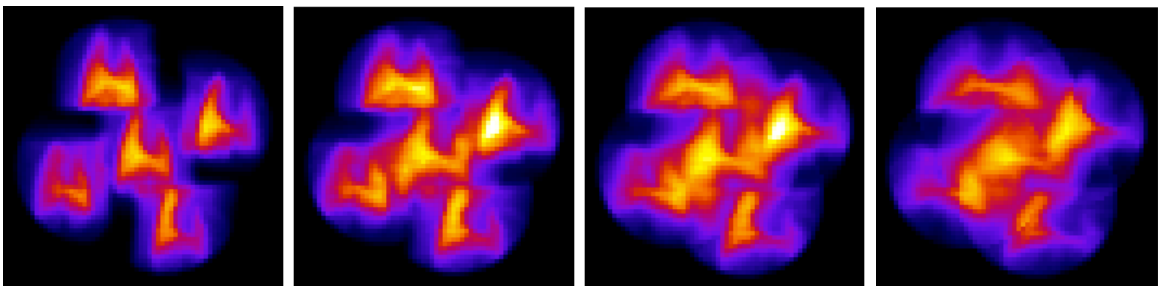
Figure 7.3: Simulated multiplexed data by decreasing the pinhole spacing and increasing the size of the lung phantoms.

Finally, an important and exciting area of investigation is the robustness of the CP method when applied to "multiplexed" data. Figure 7.2 demonstrates that if the pinholes are positioned closer to one another, or the imaged subject is larger, the pinhole projections would overlap, or multiplex, on the detector face. In multiplexed projections, there is no longer a 1 to 1 mapping from the detector coordinates to voxels in the volume for data in the regions of overlap. Since the multiplexed signal can be back-projected through multiple pinholes, this ambiguity can result in artifacts in the reconstruction. 3-view reconstructions of the data simulated above are shown in figure 7.3 using the CP method and MLEM. The results suggest the TV-penalty may be successful in breaking this ambiguity.
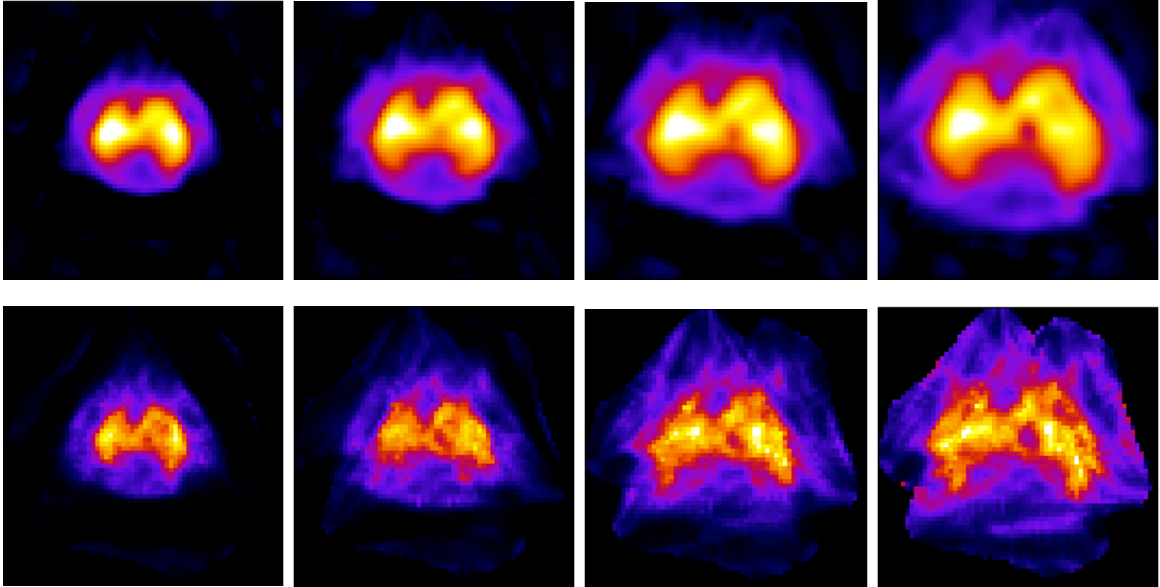
Figure 7.4: Reconstructions of the simulated multiplexed data in figure 7.2 with 5% Poisson noise. The top row are CP reconstructions using $\lambda = 1$ and $s = 100$ and the bottom row are MLEM reconstructions.

Appendix A

Gamma counting and projection image formation

Introduction

The detection and acquisition processes for the NaI gamma cameras used in this dissertation are described here. This Appendix briefly summarizes gamma counting using scintillation detectors, and describes the electronics and statistical approach used to generate a projection image from 9 PMTs.

A.1 Data acquisition and scatter rejection

The gamma camera and data acquisition electronics used on our laboratory are designed to the FastSPECT II specifications [34]. The gamma camera is connected to multi-channel pulse-height analyzers to read and digitize the output from the 9 PMTs depicted in figure 2.5. The pulse-height analyzers are connected to a large memory buffer via an ethernet cable, where data is stored in list-mode format. The electronics are then controlled using LabVIEW [55]. The pulse height analyzers, measure the height of the pulses generated from each PMT. For a single event, 9 numbers representing the pulse height of each PMT are collected in a temporary memory buffer. During acquisition a list of all events including their acquisition time and pulse height are accumulated over time then stored for post processing. Data collected in this manner is known as a "list-mode" acquisition [9, 78]. Using list-mode data, the projection images are calculated from the event list after the fact. The conversion from list-mode data to a projection image

involves two tasks; position estimation and the rejection of scatter events in the acquisition process.

Ideally, all the energy of a gamma ray photon would be converted directly to visible light, where all the visible photons are emitted along the same trajectory and location of the incident gamma photons. As illustrated in figure A.1 a gamma ray interaction with the NaI crystal results in a spread of visible photons. The energy of an incident gamma ray is therefore spread over a range of pulse heights corresponding to the range in energies of the secondary visible photons. Plotting the histogram of pulse heights
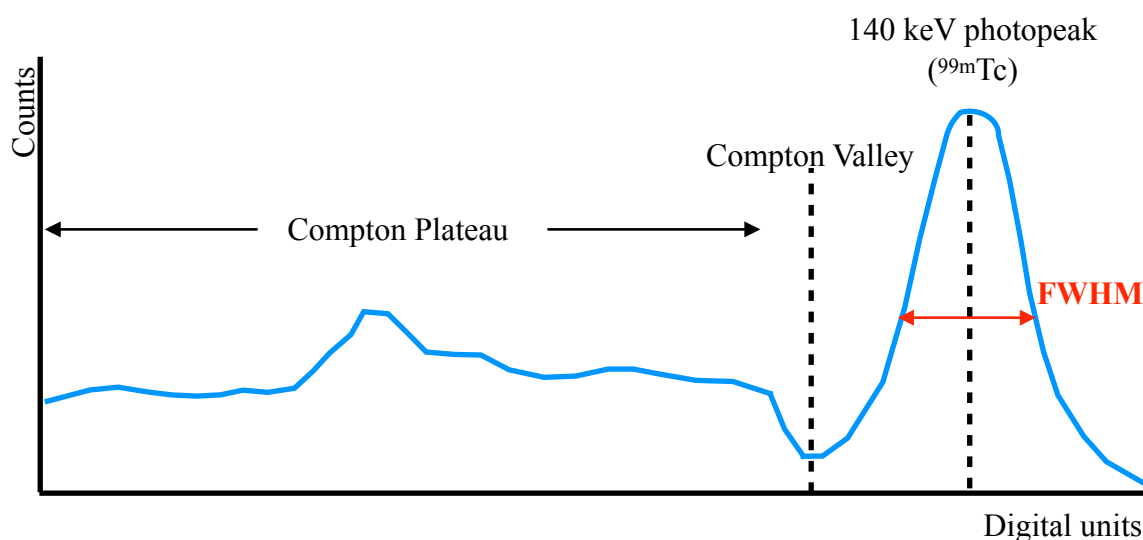
Figure A.1: Frequency spectrum of ^{99m}Tc with a photopeak corresponding to 140 keV (adapted from 14-4 in [23]).

collected from a NaI detector, a peak corresponds to the energy of the gamma emitter. The "photopeak" obtained from emissions of ^{99m}Tc correspond to an energy of 140 keV. Ideally, a 100 keV gamma ray photon will produce an average of $5000 \pm \sqrt{5000}$ visible photons in a NaI crystal [86]. The geometry of the crystal and the conversion from visible

photons to electrons in the PMT can effect how broad the photopeak is for a given NaI detector. The Full Width at Half Max (FWHM) of the photo peak is a measure of the efficiency of the detector for a given gamma emitter [86].

The path of the incident gamma ray may also be redirected through non-ideal interactions with the NaI crystal prior to conversion to visible light. These interactions are known as scatter events and make up the low energy events in the energy histogram. Scattering of gamma ray photons within the range of energies used clinically are attributed to Compton scattering and the photoelectric effect [23, 86], and these interactions are illustrated in figure A.2. In Compton scattering, a gamma ray photon interacts with a free electron in the NaI crystal. In these collisions, some energy from the incident gamma ray is transferred to the electron. To conserve momentum, the gamma photon's energy is reduced, and the trajectory is redirected. Subsequent scintillation events within the NaI crystal will result in pulse heights corresponding to an energy lower than the initial energy of the incident photon. In the photoelectric effect, all of the gamma ray's energy is transferred to a free electron within the NaI crystal. The free electron is deflected and repeatedly interacts with the NaI crystal. The deflection interactions with the crystal generate multiple low energy visible photons until the electron's kinetic energy is absorbed.

Additional features in the energy histogram can be attributed to background events. Interactions between the gamma photons and lead shielding/collimators produce x-rays that can interact with the NaI crystal [23, 86]. Gamma photons that pass through the NaI crystal can interact with other components of the detector, such as the light guide. The interactions with material outside of the NaI crystal can generate x-rays directed

Compton Scattering

Photoelectric Effect
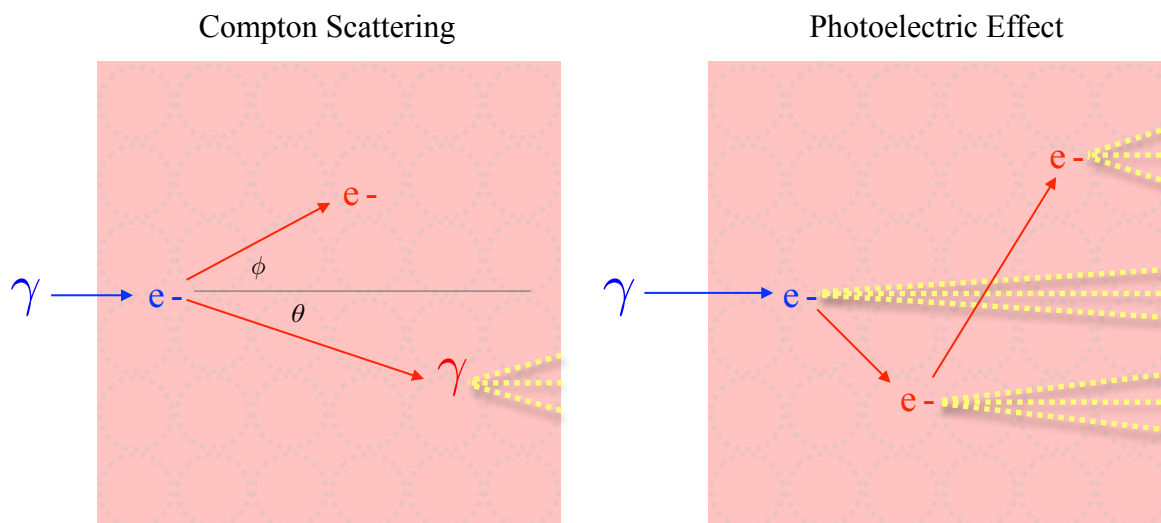
Figure A.2: Scattering in the NaI crystal: (a) Compton scattering, (b) photoelectric effect (figures adapted from [85]).

back towards the NaI crystal, generating visible photons that are reflected off the back of the detector face. These events are known as "backscatter" events [86]. Both lead shield interactions and backscatter events form characteristic photopeaks in the Compton plateau. Additionally, scattering events may occur within the subject/object being imaged, which will increase the frequency of scattering events in the detector's energy spectrum.

The list-mode pulse heights from each PMT therefore contain events due to scattering. In order to calculate a projection image from the list-mode data the event's position must be determined and events falling within the Compton plateau must be rejected, as the location of these events will not be consistent with the collimator projection model. However, rejecting events using the energy spectra, known as energy windowing [22], is complicated by the fact that the pulse heights from each PMT are not only sensitive to the energy of events but also position. Rather, as part of the camera's calibration, the response of each PMT is measured by physically placing a well-

collimated point source on a 78 x 78 grid of points spaced 0.15 mm apart on the face of the detector. For each grid point, data is acquired until 5000 events are counted. The

Figure A.3: *In situ* mean response of each of the 9 PMTs to a well collimated point source positioned on a 78 x 78 grid in front of the camera (based on figure 12.2 in [21]).

mean response for all 9 PMTs at each grid point, known as the Mean Detector Response Function (MDRF), is used to generate a look-up-table (LUT) from which events obtained during an acquisition can be compared, as pictured in figure A.3 for one of our laboratory's cameras.

The (x, y) location on the 78 x 78 grid for a given event is estimated using a maximum-likelihood position estimator and treating the MDRF as the mean estimate for all (x, y) locations [22, 9, 43]. Finally, a "likelihood window" [22, 43] procedure is used to accept or reject the event. Using the MDRF, the likelihood of the event occurring at the

estimated location with an energy of 140 keV is calculated. If the likelihood is above a fixed threshold the event is accepted. In this way, list-mode-data is sorted to a grid location or pixel. The entire projection is then the 2D grid of pixels where the value of a given pixel is the number of events assigned to its location. Using a high likelihood threshold, likelihood windowing demonstrates a similar probability of accepting 140 keV events to energy windowing, considered the gold standard, and Bayesian windowing [22].

Appendix B

Maximum Likelihood Expectation Maximization

B.1 MLEM derivation

The Maximum Likelihood Expectation Maximization (MLEM) iterative method for emission computed tomography was first proposed by Shepp and Vardi [88]. The following derivation is based on Shepp and Vardi's original paper and is shown for both completeness, as well as for consistency with this dissertation's notation.

B.2 Statistical interpretation of the forward-projection

MLEM estimation is a statistical approach to estimating a distribution of emissions from a set of projections. We begin with a statistical interpretation of our linear system in Chapter 2. Suppose we knew the activity in the jth voxel, a_j, and the contribution of voxel j^{th} to the ith detector element, a_{ij}. If we knew a_{ij}, then the problem of solving f_j would be trivial, i.e.

$$f_j = \sum_{i=0}^{M-1} a_{ij}.$$
(B.1)

Similarly, the detections counted in the ith detector would be related to a_{ij} by

$$g_i = \sum_{j=0}^{N-1} a_{ij}.$$
(B.2)

Although a_{ij} is unobservable, the expectation of B.2 may be calculated since each a_{ij} is an independent Poisson random variable. The expectations are related to the forward-

projection through h_{ij}, the probability of an emission in the jth voxel being detected in the ith detector. $E[a_{ij}]$ is given by,

$$E[a_{ij}] = p(i|j)\, a_j = h_{ij} f_j\;.$$ (B.3)

Using equation B.3, the expectation of g_i may be calculated by the following:

$$E[g_i] = E[\sum_{j=0}^{N-1} a_{ij}] = \sum_{j=0}^{N-1} E[a_{ij}] = \sum_{j=0}^{N-1} h_{ij} f_j\;.$$ (B.4)

In the limit that a sufficient number of counts are detected, we may assume that $g_i = E[g_i]$, and by definition equation B.4 is the forward-projection.

B.3 The likelihood function

The objective of MLEM is to determine f that maximizes the likelihood of obtaining g. Since a_{ij} is consistent with the forward-projection and can be related to g by equation B.4, a_{ij} can be used as a means for calculating this likelihood. We begin by calculating the probability of obtaining a_{ij} given $E[a_{ij}]$,

$$P(a_{ij}|E[a_{ij}]) = E[a_{ij}]^{a_{ij}} \left(\frac{e^{-E[a_{ij}]}}{a_{ij}!} \right) = (h_{ij} f_j)^{a_{ij}} \left(\frac{e^{-h_{ij} f_j}}{a_{ij}!} \right)$$ (B.5)

where it is assumed that a_{ij} is Poisson distributed. The probability of the observed detections in each individual detector may be obtained by summing over the voxels of equation (B.5) to obtain

$$P(g_i|f) = \sum_{j=0}^{N-1} (h_{ij} f_j)^{a_{ij}} \left(\frac{e^{-h_{ij} f_j}}{a_{ij}!} \right)$$ (B.6)

The probability of obtaining the observed data set g from the ensemble of detectors g_i is given by the likelihood function

$$L(g|f) = \prod_{i=0}^{M-1} \sum_{j=0}^{N-1} (h_{ij} f_j)^{a_{ij}} \left(\frac{e^{-h_{ij} f_j}}{a_{ij}!} \right)\;.$$ (B.7)

The product in equation B.7 can then be recast as a sum by taking the natural log of both

sides, yielding the log-likelihood,

$$log\, L(g|f) = \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} a_{ij}\, log(h_{ij}f_j) - h_{ij}f_j - log(a_{ij}!) \quad \text{(B.8)}$$

One might consider optimizing the log-likelihood to solve for the f that maximizes the

likelihood of observing the given data set g. Since a_{ij} is not known, maximizing equation

B.9 directly is not possible. Instead, we will show that a_{ij} can be eliminated by

performing an expectation on equation B.9 before maximization.

B.4 The expectation step

Taking the expectation of equation B.8 amounts to finding the conditional

expectation of a_{ij} with respect to the estimate of the reconstruction, i.e.,

$$E[logL(g|f)] = \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} E[a_{ij}|f_j]\, log(h_{ij}f_j) - h_{ij}f_j - E[log(a_{ij}!)] \quad \text{(B.9)}$$

Note that the derivative of the expectation $E[log(a_{ij}!)]$ in equation B.9 is zero, and will

be eliminated in the maximization step. It follows that the conditional distribution of a_{ij}

given the sum in equation B.2 is binomially distributed, with g_i trials, and probability p,

$$p = \frac{E[a_{ij}]}{\sum_{j} E[a_{ij}]} \quad \text{(B.10)}$$

and expectation $p\,g_i$. Therefore we obtain

$$E[a_{ij}|f_j] = p\,g_i = \frac{E[a_{ij}]}{\sum_{j} E[a_{ij}]}\, g_i = \frac{h_{ij}f_j}{\sum_{j=0}^{N-1} h_{ij}f_j}\, g_i \quad \text{(B.11)}$$

B.5 The maximization step

We may generate an estimate of the emission distribution by first maximizing (B. 11). Using basic principles of calculus, this amounts to taking the derivative with respect to f_j, and setting the derivative to zero, i.e.

$$\frac{\partial}{\partial f_j} E[logL(g|f)] = \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} E[a_{ij}|f_j] \frac{h_{ij}}{h_{ij}f_j} - h_{ij} = 0 . \quad \text{(B.12)}$$

Substituting equation B.11 in for the expectation yields,

$$\sum_{i=0}^{M-1} \frac{g_i}{\sum_k h_{ik}f_k} h_{ij} - h_{ij} = 0 , \quad \text{(B.13)}$$

where k replaces the dummy index. Equation B.13 is simplified further

$$\frac{1}{\sum_i h_{ij}} \sum_{i=0}^{M-1} h_{ji} \frac{g_i}{\sum_k h_{ik}f_k} = 1 . \quad \text{(B.14)}$$

For a given element, the image estimate is reintroduced by multiplying both sides by f_j,

$$f_j = \frac{f_j}{\sum_i h_{ij}} \sum_{i=0}^{M-1} h_{ji} \frac{g_i}{\sum_k h_{ik}f_k}, \quad \text{(B.15)}$$

which yields the final fixed point iteration proposed by Shepp and Vardi,

$$f_j^{(n+1)} = \frac{f_j^{(n)}}{\sum_i h_{ij}} \sum_{i=0}^{M-1} h_{ji} \frac{g_i}{\sum_k h_{ik}f_k} . \quad \text{(B.16)}$$

Appendix C

The Chambolle-Pock Algorithm

C.1 Introduction

This Appendix presents worked details for the iterative method proposed by Chambolle and Pock [17]. The concepts of proximal methods may be found in [11] and [63], while a brief derivation for iteratively solving TV-KL minimization problems is worked out in [92]. The following presentation is intended to summarize content from these sources, to fill in some missing details, and to supplement the discussion from chapter 4 for completeness.

C.2 Iterative Optimization And The Proximal Mapping

The proximal mapping is used by the Chambolle-Pock algorithm as a means of iteratively optimizing the primal and dual objective functions simultaneously. This section provides an interpretation of the proximal mapping as it relates to the general inversion procedure for minimizing convex functions that satisfy the first-order convexity condition.

A simple iterative procedure for minimizing a differentiable function begins with equation (C.4), where we know for a convex function, $f(x)$, the optimal value occurs where the gradient is zero. Multiplying both sides by a constant, σ, and adding x to both sides implies the iterative procedure at the bottom of equation C.1 for solving for the function's minimum. This method is the classic gradient decent method [96], illustrated

$$0 = \sigma \nabla f(x)$$

$$x = x + \sigma \nabla f(x) \tag{C.1}$$

$$x_{n+1} = x_n + \sigma \nabla f(x_n)$$

in figure (4.13) where σ is used to scale the step size in the direction of the gradient $\nabla f(x_n)$. How well the method converges to some optimum depends largely on the step size. If σ is too small, a large number of iterations are required to get a close estimate. If σ is too large, the estimates may oscillate around the optimal value. The iterative procedure in (C.1) may be rewritten as an operation performed on the current estimate x_n,

$$x_{n+1} = (\mathbf{I} + \sigma \nabla f)\, x_n, \tag{C.2}$$

where the expression in parenthesis is known as the "forward operator" [63].

For a convex function that is not smooth and continuous there is an iterative approach similar to (C.2), but there is not a well-defined forward operator, which may result from having multiple supporting hyperplanes for a given x_n. In an analogy with (C.1), we are looking for $\mathbf{0}$ in the set $\partial f(x)$. At best we can say that there is a vector x that is in the set $x + \sigma\, \partial f(x)$, which implies the following iterative procedure

$$0 \in \sigma\, \partial f(x)$$
$$x \in x + \sigma\, \partial f(x) \tag{C.3}$$
$$x_{n+1} = (\mathbf{I} + \sigma\, \partial f)^{-1} x_n \;.$$

The expression $(\mathbf{I} + \sigma\, \partial f)^{-1}$ is known as the "resolvant of the ∂f operator" [11, 17], or in some contexts known as a "backward operator" [11, 63].

The iterative procedure in equation C.3 is known as the proximal point algorithm [11, 63] for convex optimization, where the resolvant operator defines the proximal mapping

$$x_{n+1} = (\mathbf{I} + \sigma\, \partial f)^{-1} x_n = \mathrm{prox}_\sigma[f](x_n)\;. \tag{C.4}$$

Like the gradient decent method, achieving estimates close to the optimal value, x^*, depends on the choice in step size, demonstrated by Rockafeller in [83].

Since the resolvant operator is not well defined, equation C.4 requires further interpretation if the proximal mapping is to be useful for our purposes. The calculation of the proximal mapping can be shown to be the solution to the following optimization problem,

$$x = prox_\sigma[f](x') = \arg\min_{x'} \left\{ \frac{\|x - x'\|_2^2}{2\sigma} + f(x) \right\}, \tag{C.5}$$

consistent with our resolvant operator interpretation. The following is demonstrated in [11] (see proof 3.4), and shown here for completeness.

Consider $x \in (I + \sigma\partial f)^{-1} x'$, using the resolvant definition. This may be rewritten in terms of the "forward operator",

$$x' \in (I + \sigma\partial f)\, x. \tag{C.6}$$

It follows from (C.6) that

$$0 \in \frac{x - x'}{\sigma} + \partial f(x), \tag{C.7}$$

which can be rewritten in terms of the subdifferntial with respect to x',

$$0 \in \partial_{x'} \left(\frac{\|x - x'\|_2^2}{2\sigma} + f(x) \right). \tag{C.8}$$

If $f(x)$ is convex, equation C.8 is the subdiffertial form of the optimization proposed in (C.4).

C.3 Iterative Proximal Mappings For Saddle Point Problems

As mentioned in chapter 4, the proximal mappings found in the Chambolle-Pock method are designed to iteratively solve saddle point problems. Suppose we have the

functions $F^* \in Y$ and $G \in X$ and a system matrix K such that $Kx \in Y$ and $K^T y \in X$, according to the primal-dual formulation in chapter 4. If there is a solution to the optimization problem then it follows from section C.1 that there must be some optimal values x^* and a y^* that satisfy [17]

$$\sigma Kx^* \in \partial F^*(y^*)$$
$$-\tau K^T y^* \in \partial G(x^*) \, . \tag{C.9}$$

For some local x an optimization procedure may be obtained analogous to the proximal point iteration in equation C.3 is given by

$$-\tau K^T y \in \tau \, \partial G(x)$$
$$x - \tau K^T y \in x + \tau \, \partial G(x) \tag{C.10}$$
$$x_{n+1} = (I + \tau \, \partial G)^{-1} (x_n - \tau K^T y_n) \, .$$

Likewise for some local y we obtain the optimization procedure

$$\sigma Kx \in \sigma \, \partial F^*(y)$$
$$y + \sigma Kx \in y + \sigma \, \partial F^*(y) \tag{C.11}$$
$$y_{n+1} = (I + \sigma \, \partial F^*)^{-1} (y_n + \sigma Kx_n) \, .$$

Equations C.10 and C.11 are lines 4 and 5, respectfully, in algorithm 4.1.

C.4 The SPECT Primal Function And Its Dual Formulation

This section presents the convex conjugate of the constrained SPECT objective problem, i.e. D_{KL} minimization subject to TV penalization. As summarized in Chapter 4, the Chambolle-Pock method optimizes a primal objective formulated by the sum of two convex functions, F and G, where F performs a mapping onto $\mathbb{R}^M$ and G performs the mapping onto $\mathbb{R}^N$.

The function F may be split into two components, one for the forward-projection and the other for the gradient magnitude operations

$$F = F_1(\mathbf{H}f) + F_2(\nabla f) = D_{KL}(\mathbf{H}f, g) + (\lambda/\nu)\| |\nu\nabla f| \|_1 .$$ (C.12)

We begin by presenting the convex conjugate of the forward-projection component of F, or F_1,

$$F_1(\mathbf{H}f) = F_1(a) = \sum_i a - g + g\,ln(g/a) .$$ (C.13)

From our objective function, F_1 is the D_{KL} data divergence, where the vector a substitutes for $\mathbf{H}f$. The convex conjugate is obtained by taking the Legendre transform of F_1,

$$F_1^*(y) = \max_a \{a^T y - F_1(a)\} ,$$ (C.14)

defined by F_1^*, where the vectors a and y form a "conjugate pair" [12]. If we let a^* be the vector that maximizes C.14, then the expression is maximized when the following criteria are met:

$$\partial_a(a^T y - F_1(a)) = 0$$

$$y - \partial_a(F_1(a)) = 0$$

$$y - \left\{ \mathbf{1} + g\left|\frac{a}{g}\right| \cdot \left(\frac{-g}{a^2}\right) \right\} = 0$$ (C.15)

$$y - \left(\mathbf{1} - \frac{g}{a}\right) = 0$$

$$a = \frac{g}{\mathbf{1}-y} = a^* .$$

Using a^* as the maximizer of equation C.14, the convex conjugate of F_1 becomes

Simplifying equation C.16, yields the final conjugate function,

$$F_1^*(y) = \left(\frac{g}{\mathbf{1}-y}\right)^T y - \left[\sum_i \frac{g}{\mathbf{1}-y} - g + g\,ln(\mathbf{1}-y)\right] ,$$ (C.16)

The convex conjugate of F_2 may be taken using the Legendre transformation, as before. For the sake of notation, b will be used in place of $|\nu \nabla f|$ and μ will be used in place of λ/ν. The Legendre transform is then given by

$$F_1^*(y) = \sum_i -g\,ln(\mathbf{1} - y) \tag{C.17}$$

where z and b are our conjugate pairs. Evaluating the Legendre transformation is not as straightforward as before, since the total-variation is not differentiable.

$$F_2^*(z) = \max_b \left\{ b^T z - \mu\|b\|_1 \right\}, \quad \text{for } \mu \leq 1, \tag{C.18}$$

In the Legendre transformation of F_2 we are comparing how large $b^T z$ is relative to $\|b\|_1$. This can be found by taking their ratio,

$$\frac{b^T z}{\|b\|_1}. \tag{C.19}$$

We may further consider the scenario when this ratio is its largest possible value, which is defined by the dual norm of z,[1]

$$\|z\|_* = \max_{b \neq \mathbf{0}} \left\{ \frac{b^T z}{\|b\|_1} \right\}. \tag{C.20}$$

In general, this ratio, and subsequently the dual norm, is either less than or equal to 1 or strictly greater than 1.

In the case where $\|z\|_* \leq 1$, the dual norm will always be greater than or equal to the ratio in equation C19. Hence we get

$$b^T z \leq \|b\|_1 \|z\|_* \quad \forall b, \tag{C.21}$$

similar to the Cauchy-Schwartz inequality. It follows that if $\mu \leq 1$, then the difference in the Legendre transform will always be negative or zero,

$$b^T z - \mu\|b\|_1 \leq 0, \tag{C.22}$$

---

[1] The dual-norm of the 1-norm, $\|b\|_1$, is the maximum value in the vector, or the "sup-norm," $\|b\|_\infty$

and is therefore maximal when zero.

Similarly, in the case where $\|z\|_* > 1$, it follows from the ratio in equation C.19 that there is a b such that $\|b\|_1 \leq 1$ and $b^T z > 1$. For $\mu \leq 1$, the difference in the Legendre transform is always greater than zero,

$$b^T z - \mu\|b\|_1 > 0. \tag{C.23}$$

The difference is then maximal when the difference approaches infinity,

$$\{b^T z - \mu\|b\|_1\} \to \infty. \tag{C.24}$$

Combining these two possibilities, the Legendre transform $F_2^*(z)$ is then given by the indicator function,

$$F_2^*(z) = \begin{cases} 0, & \|z\|_* \leq 1 \\ \infty & \|z\|_* > 1 \end{cases} = \delta(\|z\|_*) = \delta_\mu(|z|). \tag{C.25}$$

As mentioned above, the function G does not explicitly appear in the objective function, and hence $G(x) = \mathbf{0}$. Functionally, however, the argument of G must perform a mapping onto $\mathbb{R}^N$, where N is the number of voxels. We may define the argument x by

$$x = \mathbf{H}^T y + \nu \nabla^T \vec{z}, \tag{C.26}$$

where the transpose operations map y and z into $\mathbb{R}^N$, which we call the *dual condition*. Our solution to the objective function will only be realized as these two components begin to cancel out. The Legendre transform of $G(x)$ is then given by

Equation C.27 is maximized for $q = \mathbf{0}$, or for very large q. So the Legendre transform (C.27)

$$G^*(q) = \max_x \{x^T q - \mathbf{0}\}. $$

$G^*(q)$ is also an indicator function,

$$G^*(q) = \begin{cases} 0, & q = \mathbf{0} \\ \infty & q \neq \mathbf{0} \end{cases}. \tag{C.28}$$

C.5 The Proximal Mappings For TV-KL minimization

The iterative procedure for the primal problem can be formed by taking the proximal mapping of $G(x)$. Since $G(x) = 0$, the proximal mapping is given as,

$$f = prox_\tau[G](x) = \arg\min_f \left(\frac{\|x - f\|_2^2}{2\tau} \right) . \qquad (C.29)$$

The optimum to equation C.29 is

$$f = x = \mathbf{H}^T y + \nu \nabla^T \vec{z}' , \qquad (C.30)$$

where we substitute the dual condition for x. The saddle point iteration, from equation C. 10, then yields line 11 of algorithm 4.2,

$$f_{n+1} = f_n - \tau \left(\mathbf{H}^T y + \nu \nabla^T \vec{z}' \right) . \qquad (C.31)$$

Using the conjugate function F^*, we can form the proximal mapping for the dual problem. Evaluation of the proximal mapping of F^* can be simplified by exploiting the separable sum property of the proximal operator. Namely, if the function is separable in two variables, $F^*(y, \vec{z}) = F_1^*(y) + F_2^*(\vec{z})$, then the proximal mapping may be evaluated as,

$$prox_\sigma[F^*](y, \vec{z}) = (prox_\sigma[F_1^*](y), prox_\sigma[F_2^*](\vec{z})) . \qquad (C.32)$$

Thus, the proximal mappings for $F_1^*(y)$ and $F_2^*(\vec{z})$ can be evaluated independently [11]. For $F_1^*(y)$, $\qquad (C.33)$

$$prox_\sigma[F_1^*](y) = \arg\min_{y'} \left\{ F_1^*(y') + \frac{\|y - y'\|_2^2}{2\sigma} \right\} = \arg\min_{y'} \left\{ \sum_i [-g\,ln(\mathbf{1} - y')]_i + \frac{\|y - y'\|_2^2}{2\sigma} \right\} .$$

In its differential form, equatio C.33 may be written as

$$\partial_{y'} \left\{ \sum_i [-g\,ln(\mathbf{1} - y')]_i + \frac{\|y - y'\|_2^2}{2\sigma} \right\} = 0 , \qquad (C.34)$$

where the optimal solution is obtained by differentiating with respect to y', and solving

the subsequent quadratic equation for y',

$$y' = prox_\sigma[F_1^*](y) = \tfrac{1}{2}\left((\mathbf{1}+y) \pm \sqrt{(\mathbf{1}-y)^2 + 4\sigma g}\right). \qquad \text{(C.35)}$$

The saddle point iteration, from equation C.11, is recovered by letting

$y_{n+1} = y_n + \sigma \mathbf{H}\bar{f}_n$, which yields line 4 of algorithm 4.2.

To complete the dual problem, the proximal mapping for $F_2^*(\vec{z})$ is computed by

the following, (C.36)

$$prox_\sigma[F_2^*](\vec{z}) = \arg\min_{\vec{z}'} \left\{ F_2^*(\vec{z}') + \tfrac{\|\vec{z}-\vec{z}'\|_2^2}{2\sigma} \right\} = \arg\min_{\vec{z}'} \left(\delta_\mu + \tfrac{\|\vec{z}-\vec{z}'\|_2^2}{2\sigma} \right), \ \|z\|_* \leq 1.$$

In its differential form, (C.36) may be written as,

$$\partial_{\vec{z}_i'}\left(\tfrac{|\vec{z}_i - \vec{z}_i'|^2}{2\sigma}\right) = 0, \ \ \forall i \ , \qquad \text{(C.37)}$$

where the optimal solution is obtained by differentiating with respect to $\vec{z}'$ and solving

the for $\vec{z}'$,

$$\vec{z}' = \vec{z} \ . \qquad \text{(C.38)}$$

As before, the saddle point iteration, from equation C.11, is obtained by letting

$\vec{z}_{n+1} = \vec{z}_n + \sigma\nu\nabla\bar{f}_n$, which yields line 8 in algorithm 4.2.

Equation C.36 is constrained by the indicator function, $\delta_\mu(|\vec{z}|)$ in C.25, such that

$\|z\|_* \leq 1$. This is enforced on line 9 of algorithm 4.2,

$$\vec{z}_{n+1} = \vec{z}_n' \frac{\mu}{max(\mu, |\vec{z}_n'|)} = \vec{z}_n' \frac{\lambda/\nu}{max(\lambda/\nu, |\vec{z}_n'|)} \ . \qquad \text{(C.39)}$$

C.6 Additional Constraints

Minimizing the TV-penalized D_{KL} objective function (equation 4.25) requires some practical constraints for a robust implementation of the CP method. Ideally, pseudo data obtained from the exact solution to our optimization problem should agree with the projection data. That is,

$$\|\mathbf{H}f - g\|_2 = 0,\tag{C.40}$$

when the exact solution is reached. For noisy data, this requirement is too restrictive, as the exact solution cannot be achieved. At best, after many iterations of the CP method, the difference between the data, g, and the pseudo data $\mathbf{H}f$ should be small. A more realistic constraint would require the method to iterate until the difference between the data and pseudo data is within some small number,

$$\|\mathbf{H}f - g\|_2 \leq \epsilon,\tag{C.41}$$

where ϵ is the data-error tolerance.

The data-error tolerance may be incorporated into the objective function by restricting our estimates to the set of solutions where equation C.41 is true. The solution sets of interest can be defined in terms of the 2-norm ball [91],

$$B_2(\epsilon) = \{\mathbf{H}f - g \mid \|\mathbf{H}f - g\|_2 \leq \epsilon\}.\tag{C.42}$$

We can then incorporate the constraint into our objective function using the following indicator function,

$$\delta_{B_2(\epsilon)}(\mathbf{H}f - g) = \begin{cases} 0, & \mathbf{H}f - g \in B_2(\epsilon) \\ \infty & \mathbf{H}f - g \notin B_2(\epsilon). \end{cases}\tag{C.43}$$

If we let $a = \mathbf{H}f$, and $F(a) = \delta_{B_2(\epsilon)}(a - g)$, the conjugate $F^*(y)$ is formed using the Legendre transform as

$$\tag{C.44}$$

$$F^*(y) = \max_a \left\{ a^T y - F(a) \right\} = y^T g + \max_{a-g} \left\{ (a - g)^T y - \delta_{B_2(\epsilon)}(a - g) \right\}.$$

For $a \in B_2(\epsilon)$, equation C.44 is maximized when $(a - g) = \epsilon\, y/\|y\|_2$ [91]. The conjugate is then,

$$F^*(y) = y^T g + \epsilon\|y\|_2$$

(C.45)

The proximal mapping of equation C.45 then follows from our definition in C.5,

$$y' = prox_\sigma[F](y) = \arg\min_{y'} \left\{ \frac{\|y - y'\|_2^2}{2\sigma} + y'^T g + \epsilon y' \right\}$$

(C.46)

The proximal mapping in equation C.46 is demonstrated in [92], using geometric arguments, and briefly in [91]. The result, in an iterative context, leads to line 7 in algorithm 4.2.

We may also consider constraining the intermediate y' values in C.35 to be positive. Although an additional indicator function may be used to enforce positivity, it amounts to choosing the negative radical in equation C.35 [92].

Appendix D

Experimental Details

D.1 Introduction

This Appendix discusses details from chapter 6. This includes an alternative approach to obtaining the system matrix using strategy (2) in 6.3.3. Additionally, the threshold procedure used to segment the heart/liver regions from section 6.5.1 is presented.

D.2 Re-weighting the pinhole projections

This section briefly describes how the normalization strategy (2), described in chapter 6.3, can be modified to better match the projection data. The pseudo projections using the normalization (2) may be improved by re-weighting the contribution of each pinhole to match the experimental data. In normalization procedure (2) each pinhole's projection is normalized to 1. However, in experimental data the object may sit slightly higher or lower than the center line of the projection image. The projection from the pinhole closest to the object will therefore be brighter than the projection from the pinhole furthest from the object. Therefore, each pinhole's response should be weighted according to the mean intensity of each pinhole's projection, rather than normalizing all the pinholes to 1. In order to obtain the response from each pinhole, a circular region of interest may be placed over the corresponding projection and the mean pixel activity can be calculated in that region over all view angles. After normalizing each pinhole

separately, in procedure (2), the pinhole projections may then be weighted relative to the maximum mean response from the data. The procedure for re-weighting each pinhole's response is illustrated in figure D.1. In general, each individual projection may deviate

(a)

(b)
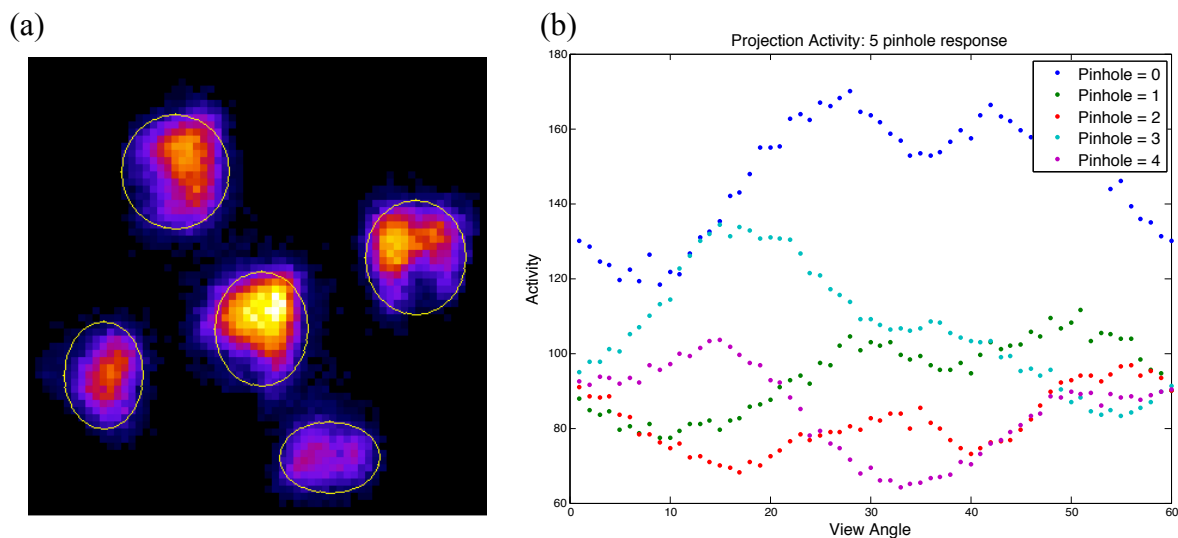
Figure D.1: Procedure for determining the mean response of each pinhole: (a) masking each pinhole projection, (b) the response of each pinhole per view angle.

slightly from this mean response due to the animal's orientation, making the approach proposed in procedure (2) less adaptive.

MAA reconstructions using strategy (2) are shown in figure D.2 for a range of injected activities. Compared to the 3 view reconstructions figure 6.14 and 6.19 the reconstructions, there is less resolution between the lung regions and the background. This is particularly evident for the 3-view 0.2 mCi reconstruction, where the shape of the lungs cannot be made out. Additionally, the streak artifacts are distributed over a larger region of voxels outside the rat. Future work may investigate a different normalization procedure using strategy (2), that is, normalizing the projections from each pinhole separately. Figure 6.10 demonstrates that this matrix model cannot sense when the object
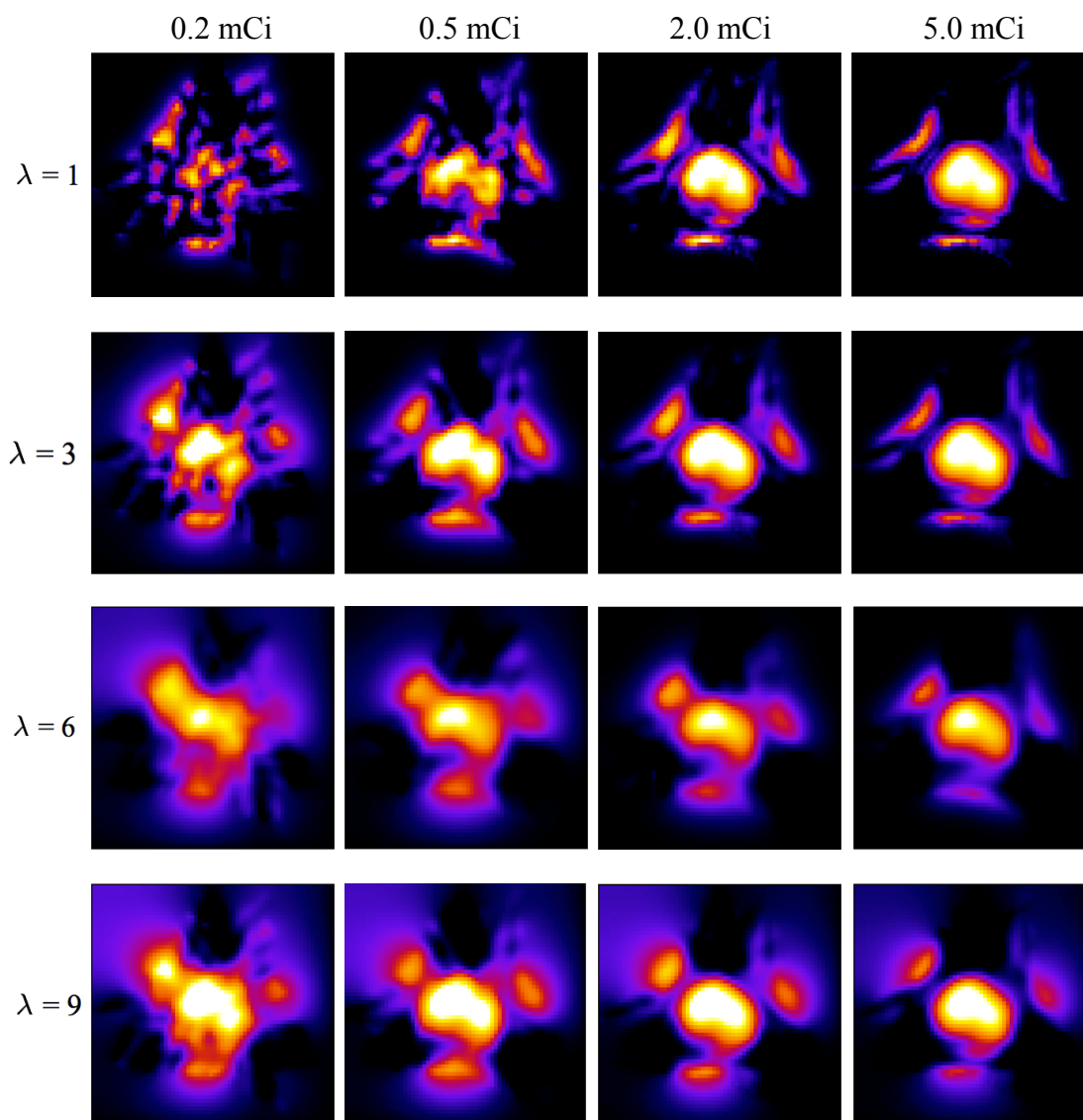
Figure D.2: 3-view reconstructions of ^{99m}Tc-MAA data with activity between 0.2 mCi and 5.0 mCi. The smoothing parameter λ is adjusted for each reconstruction, with $s = 1$.

is closer to one pinhole over the other, as indicated by the equally weighted pinhole projections. This suggests the $1/r^2$ weighting may get normalized out of the projection model in the current form of strategy (2), but not in strategy (1).

D.3 Threshold procedure for obtaining the heart and liver regions in the ⁹⁹ᵐTc-HMPAO reconstructions

The threshold procedure for segmenting out the heart/liver regions from the 60 view angle HMPAO reconstructions was performed by picking a range of thresholds to test the number of voxels in the volume after the threshold was applied.

First, Otsu's method was chosen, since this a widely used threshold level, and was calculated for the entire HMPAO reconstruction volume. Then a range of thresholds were generated above and below Otsu's level. To do this, thresholds centered about the Otsu level were then generated such that

$$t \in [2 \cdot Otsu - max, \, max] \tag{D.1}$$

where *max* is the maximum intensity in the reconstruction volume. For each *t*, a mask of the voxels with intensities greater than *t* are formed for a given HMPAO volume, as in figure D.3. The number of nonzero voxels in each mask was then determined and plotted

Figure D.3: Masks created by thresholding an HMPAO reconstruction with *t* increasing from left to right. White represents voxel intensities greater than *t*.

over *t* in figure   D.4. The mean number of voxels at each *t* was calculated from all HMPAO reconstructions in figure D.4 to keep the figure from being cluttered. The variation in the voxel count at each *t* is primarily due to the variation in intensities in the HMPAO reconstruction (due to the uptake or lack thereof), and the slight variation in size
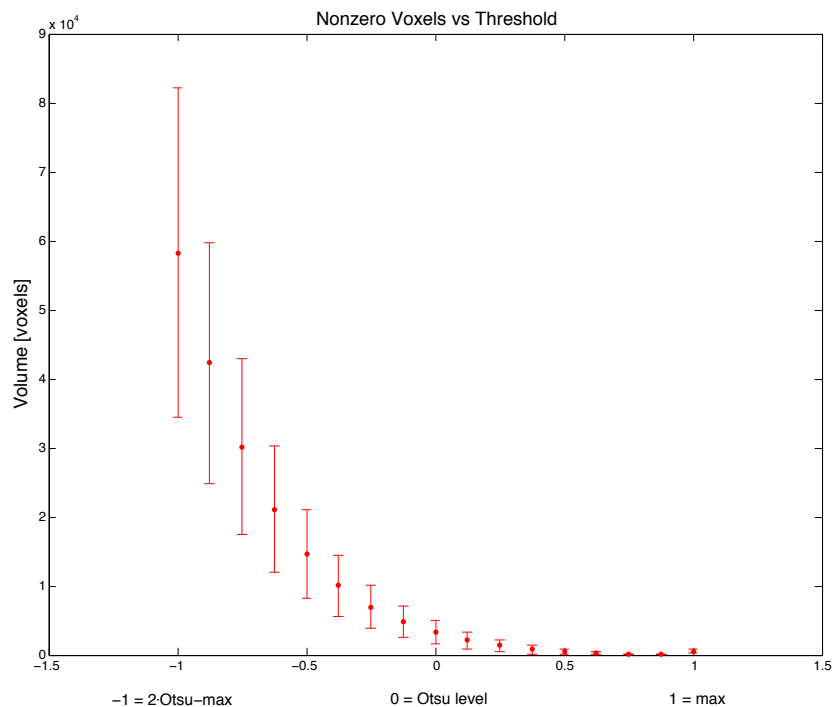
Figure D.4: The mean number of voxels leftover after thresholding above and below Otsu's level. The mean and standard deviation was calculated over all HMPAO reconstructions.

of the animals. The *t* corresponding to regions of the HMPAO reconstructions with the most uptake is therefore different for each volume.

However, as the threshold level is increased, the number of voxels with intensities above the threshold level decays to zero for all reconstructions. For each HMPAO reconstruction, the mean number of voxels from the set of masks, as in the set of masks from figure D.3, was calculated. The *t* mask with nonzero voxels closest to this mean was chosen as the heart mask.

BIBLIOGRAPHY

[1] Abràmoff, M. D., Magalhães, P. J., & Ram, S. J. (2004). Image processing with ImageJ. *Biophotonics International, 11*(7), 36-42.

[2] Alhassen, F., Kim, S., Sayre, G.A., Bowen, J.D., Gould, R., Seo, Y., Kudrolli, H., Singh, B., Nagarkar, V.V. (2011). Ultrafast Multipinhole Single Photon Emission Computed Tomography Iterative Reconstruction Using CUDA. *IEEE Nuclear Science Symposium Conference,* San Jose, CA, USA, MIC9.S-181.

[3] Anger, Hal O. (1958). Scintillation Camera. *Review of Scientific Instruments 29*, 27. doi:10.1063/1.1715998

[4] Asenbaum, S., Brücke, T., Pirker, W., Pietrzyk, U., Podreka, I. (1998) Imaging of cerebral blood flow with technetium-99m-HMPAO and technetium-99m-ECD: a comparison. *Journal of Nuclear Medicine, 39*(4), 613-618.

[5] Audi, S. H., Jacobs, E.R., Zhang, X., Camara, A.K., Zhao, M., Medhora, M.M., Rizzo, B.M., Clough, A.V. (2017). Protection by Inhaled Hydrogen Therapy in a Rat Model of Acute Lung Injury can be Tracked in vivo Using Molecular Imaging. *Shock, 48*(4), 467-476.

[6] Audi, S. H., Roerig, D. L., Haworth, S. T., Clough, A.V. (2012). Role of Glutathione in Lung Retention of 99mTc-Hexamethylpropyleneamine Oxime in Two Unique Rat Models of Hyperoxic Lung Injury. *Journal of Applied Physiology, 113*(4)*,* 658-665.

[7] Barrett, H. H., Furenlid2, L. R., Freed, M., Hesterman, J. Y., Kupinski, M. A., Clarkson, E., Whitaker, M. K. (2008). Adaptive SPECT. *IEEE Trans Med Imaging, 27*(6), 775–788. doi:10.1109/TMI.2007.913241

[8] Barrett, H. H., & Myers, K. J. (2004). *Foundations of image science.* Hoboken, New Jersey: Wiley series in pure and applied optics: John Wiley and Sons.

[9] Barrett, H. H., White, T., Parra, L.C. (1997). List-mode Likelihood. *Journal of the Optical Society of America Optics, Image Science, and Vision. 14*(11): 2914–2923. doi: 10.1364/JOSAA.14.002914

[10] Beekman, F. J., Vastenhouw, B. (2004). Design and simulation of a high-resolution stationary SPECT system for small animals. *Physics in Medicine and Biology, 49*(19)*,* 4579–4592.

[11] Boyd, S. P., Parikh, N. (2013). Proximal Algorithms. *Foundations and Trends in Optimization, 1*(3) 123-231.

[12] Boyd, S. P., Vandenberghe, L. (2009). *Convex Optimization*. New York, NY: Cambridge University Press.

[13] Bruyant, P. P. (2002). Analytic and iterative reconstruction algorithms in SPECT. *The Journal of Nuclear Medicine, 43*(10), 1343–1358.

[14] Bushberg, J.T. (2002). Lippincott Williams & Wilkins, *Essential physics of medical imaging (2nd Edition)*. Philadelphia, Pennsylvania.

[15] Candes, E., Romberg, J., & Tao, T. (2006). Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics, 59*(8), 1207–1223. doi:10.1002/cpa.20124

[16] CERETEC™, *GE Healthcare*, Chicago, IL.

[17] Chambolle, A., & Pock, T. (2011). A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision, 40*(1), 120-145. doi:10.1007/s10851-010-0251-1

[18] Chávez, C. E., Alonzo-Atienza, F., Àlvarez, D. (2013). Avoiding the Inverse Crime in the Inverse Problem of Electrocardiography: Estimating the Shape and Location of Cardiac Ischemia. *Computing in Cardiology, 40*, 687-690.

[19] Chen J.l., Zhang H.m., Ya B., Lie L., Guan M., Wang L.y. (2013) Matrix Approach for Processing of Iterative Reconstruction on Cone Beam CT. *IEEE, Conference on Medical Imaging Physics and Engineering (ICMIPE)*, Shenyang, China. doi: 10.1109/ICMIPE.2013.6864506

[20] Chen, K., Chen, X. (2010). Design and development of molecular imaging probes. *Current Topics in Medicinal Chemistry, 10*(12), 1227–1236.

[21] Chen, Y.C., Furenlid, L.R., Wilson, D.W., Barrett, H.H. (2005). Small-Animal SPECT Imaging, ch. 12. *Springer*. 195-201.

[22] Chen, J.C., (1997). Scatter rejection in modular gamma cameras for use in dynamic 3d spect brain imaging system. *Computerized Medical Imaging and Graphics, 21*(5), 283-291.

[23] Cherry, S. A., Sorenson, J.A., Phelps, M.E. (2003). Saunders, *Physics in nuclear medicine*, (3rd ed.). Philadelphia, Pennsylvania: Saunders.

[24] Clough, A. V., Audi, S. H., Haworth, S. T., & Roerig, D. L. (2012). Differential lung uptake of 99mTc- hexamethylpropyleneamine oxime and 99mTc-duramycin in the chronic hyperoxia rat model. *Journal of Nuclear Medicine, 53*(12), 1984-1991. doi:10.2967/jnumed.112.108498

[25] Clough, A. V., Barrett, H. H. (1983). Attenuated Radon and Abel transforms. *Journal of the Optical Society of America, 73*(11), 1590-1595.

[26] Cooley, J. W., Tukey, J. W. (1965). An Algorithm for the Machine Calculation of Complex Fourier Series. *Mathematics of Computation, 19*(90) 297-301.

[27] Cormack, A.M. (1963). Representation of a Function by Its Line Integrals, with Some Radiological Applications. *Journal of Applied Physics, 34*(9), 2722-2727. doi:10.1063/1.1729798

[28] Crapo, J.D., Barry, B.E., Foscue, H.A., Shelburne, J. (1980). Structural and biochemical changes in rat lungs occurring during exposures to lethal and adaptive doses of oxygen. *American Review of Respiratory Disease*, *122*(1), 123-143.

[29] Dothager, R.S., Piwnica-Worms D. (2009). Molecular imaging of pulmonary disease in vivo. *Proceedings of The American Thoracic Society, 6* 403-410.

[30] DRAXIMAGE®. DRAXIMAGE MAA, https://www.accessdata.fda.gov/ drugsatfda_docs/label/2009/017881s010lbl.pdf

[31] Exxact Corporation, Fremont, CA.

[32] Feldkamp, L. A., Davis, L. C., & Kress, J. W. (1984). Practical cone-beam algorithm. *Journal of the Optical Society of America A: Optics, Image Science, and Vision, 1*(6), 612-619. doi:10.1364/JOSAA.1.000612

[33] Foucart, S., Rauhut, H. (2013). *A Mathematical Introduction to Compressive Sensing*. New York, NY: Springer New York.

[34] Furenlid, L. R., Wilson, D. W., Chen, Y., Kim, H., Pietraski, P. J., Crawford, M.J., Barrett, H. H. (2004). FastSPECT II: A Second-Generation High-Resolution Dynamic SPECT Imager. *IEE Transactionactions on Nulcear Science, 51*(3), 631-635. doi:10.1109/TNS.2004.830975

[35] Galvin, I., Drummond, G.B., Nirmalan, M. (2007). Distribution of blood flow and ventilation in the lung: gravity is not the only factor. *British Journal of Anaesthesia, 98*(4), 420-428. doi:10.1093/bja/aem036

[36] Gajic O., Dara S.I., Mendez J.L., Adesanya A.O., Festic E., Caples S.M., Rana R., St Sauver J.L., Lymp J.F., Afessa B., Hubmayr R.D. (2004). Ventilator-associated lung injury in patients without acute lung injury at the onset of mechanical ventilation. *Critical Care Medicine 32*(9), 1817-1824.

[37] Gilbert, A., Indyk, P. (2010). Sparse Recovery Using Sparse Matrices. *Proceedings of the IEEE, 98*(6), 937-947. doi:10.1109/JPROC.2010.2045092

[38] Gonzalez, R. C., & Wood, R. E. (2008). In Pearson Prentice Hall, *Digital image processing (3rd Edition)*. Upper Saddle River, New Jersey: Pearson Prentice Hall.

[39] Gordon, R., Bender, R., Herman, G. T. (1970). Algebraic Reconstruction Techniques (ART) for three-dimensional electron microscopy and X-ray photography. *Journal of Theoretical Biology, 29*(3), 471-476.

[40] Han, G., Liang, Z., & You, J. (1999). A fast ray-tracing technique for TCT and ECT studies. *IEEE Nuclear Science Symposium 3, Seattle, WA, USA*. 1515-1518. doi: 10.1109/NSSMIC.1999.842846

[41] Haupt, J., Nowak, R. (2006). Signal Reconstruction From Noisy Random Projections. *IEEE Transactions on Information Theory, 52*(9), 4036-4048. doi: 0.1109/TIT.2006.880031

[42] Herasevich V., Yilmaz M., Khan H., Hubmayr R.D., Gajic O. (2009). Validation of an electronic surveillance system for acute lung injury. *Intensive Care Medicine 35*(6), 1018-1023.

[43] Hesterman, J.Y., Caucci, L., Kupinski, M. A., Barrett, H. H., Furenlid, L. R. (2010). Maximum-Likelihood Estimation With a Contracting-Grid Search Algorithm. *IEEE Trans Nucl Sci. 2010 June 1; 57*(3): 1077–1084. doi:10.1109/TNS. 2010.2045898

[44] Hill, D. L., Batchelor, P. G., Holden, M., Hawkes, D. J., Medical image registration. (2001). *Physics in Medicine and Biology 46*(3), R1–45.

[45] Honiden S., Gong M.N. (2009). Diabetes, insulin, and development of acute lung injury. *Critical Care Medicine 37*(8), 2455-2464.

[46] Hounsfield, G. Computerized transverse axial scanning (tomography). (1973). *British Journal of Radiology. 46*, 1016-1022.

[47] Hudson, M. H., & Larkin, R. S. (1994). Accelerated image reconstruction using ordered subsets of projection data. *IEEE Transactions on Medical Imaging, 13*(4), 601-609. doi:10.1.1.21.5810

[48] Hui, P., Haoran, C., Debasis, M., Grant, T.G., Youngho, S. (2017) Sparse domain approaches in dynamic SPECT imaging with high-performance computing. *American Journal of Nuclear Medicine and Molecular Imaging. 7*(6), 283-294.

[49] Intel ®, *Intel*, Santa Clara, Ca.

[50] Jha, A. K., Song, Na. Caffo, B. Frey, E. C. (2015). Objective evaluation of reconstruction methods for quantitative SPECT imaging in the absence of ground truth. *Proceedings of the SPIE: Medical Imaging 2015: Image Perception, Observer Performance, and Technology Assessment,* Orlando, FL, USA. *, 9461.* doi: 10.1117/12.2081286

[51] Johnson, R., Hu, H., Haworth, S. T., Cho, P. S., Dawson, C. A., & Linehan, J. H. (1998). Feldkamp and circle-and-line cone-beam reconstruction for 3D micro-CT

of vascular networks. *Physics in Medicine and Biology, 43*(4), 929-940. doi: 10.1088/0031-9155/43/4/020

[52] Jørgensen, J.S., Sidky, E.Y. (2015). How little data is enough? Phase-diagram analysis of sparsity-regularized X-ray computed tomography. *Philosophical Transactions of the Royal Society A, 373*(2043). doi:10.1098/rsta.2014.0387

[53] Kak, A. C., Slaney, M. (1988). *Principles of Computerized Tomographic Imaging*. New York, NY: IEEE Press.

[54] Kupinski, M. A., Barrett, H. H. (2005). Springer Science, *Small-Animal SPECT Imaging*. Springer, New York, New York.

[55] LabVIEW. *National Instruments*, Austin, TX.

[56] Lalush, D. S. (2003). *Emission Tomography: The Fundamentals of PET and SPECT*. Iterative Reconstruction Algorithms. San Diego, CA.: Academic Press

[57] Lalush, D. S., Tsui, B. M. W. (1994). Improving the convergence of iterative filtered backprojection algorithm. *Medical Physics, 21*(8), 1283–1286.

[58] Landsman, W. (1992). NASA IDL Astronomy Library. *https://idlastro.gsfc.nasa.gov/ftp/pro/math/poidev.pro*. Retrieved July, 2014.

[59] Lassen, N. A., Andersen, A. R., Friberg, L., & Paulson, O. B. (1988). The retention of [99mTc]-k, l-HM- PAO in the human brain after intracarotid bolus injection: A kinetic analysis. *Journal of Cerebral Blood Flow & Metabolism, 8*, S13–S22. doi: 10.1038/jcbfm.1988.28

[60] Leonard, J., Nowotnik, D. P., & Neirinckx, R. D. (1986). Technetium-99m d,l-HM-PAO: A new radiopharmaceutical for imaging regional brain perfusion using SPECT. A comparison with iodine-123 HIPDM. *Journal of Nuclear Medicine, 27*(12), 1819-1923.

[61] Levitzky, M.G. (1982). *Pulmonary Physiology*. McGraw-Hill, New York, New York.

[62] Liang, Z., Turkington, T. G., Gilland, D. R., Jaszczak, R. J., & Coleman, R. E. (1992). Simultaneous compensation for attenuation, scatter and detector response for SPECT reconstruction in three dimensions. *Physics in Medicine and Biology, 37*(3), 587-603. doi:10.1088/0031-9155/37/3/007

[63] Lorenz, D. A., Pock, T. (2015). An Inertial Forward-Backward Algorithm for Monotone Inclusions. *Journal of Mathematical Imaging and Vision. 51*(2)*, 311-325. doi: 10.1007/s10851-014-0523-2

[64] Lustig, M., Donoho, D., & Pauly, J. M. (2007). Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magnetic Resonance in Medicine, 58*(6), 1182–1195. doi:10.1002/mrm.21391

[65] Ma, Dan. Clough, Anne V., Gilat-Schmidt, Tally. (2010) Multi-pinhole dynamic SPECT imaging: simulation and system optimization. Proc. *SPIE 7622, Medical Imaging 2010: Physics of Medical Imaging, 76220U*. doi:10.1117/12.844106

[66] MATLAB ® (2013b). *The MathWorks,* Natick, MA.

[67] Matthay, M.A., Zemans R.L. (2011). The acute respiratory distress syndrome: pathogenesis and treatment. *Annu Rev Pathol 6,* 147-163

[68] McLachlan, G.J., Thriyambakam, K. (2008). Wiley Inter-Science, *The EM Algorithm and Extensions (2nd Edition)*. Hoboken, New Jersey: John Wiley and Sons, Inc.

[69] MPI. Open MPI: Open Source High Performance Computing. Retrieved 6/10, 2014, from https://www.open-mpi.org

[70] Mukherjee J.M., Dey, J., King, M.A., Konate, S. (2013). Objective assessment of penalized maximum likelihood reconstruction with sparsity-promoting penalty for myocardial perfusion SPECT imaging. *Proceedings of the SPIE: Physics of Medical Imaging*, Lake Buena Vista, FL, USA., *8668*

[71] Neirinckx, R. D., Burke, J. F., Harrison, R. C., Forster, A. M., Andersen, A. R., & Lassen, N. A. (1988). The retention mechanism of technetium-99m-HM-PAO: Intracellular reaction with glutathione. *Journal of Cerebral Blood Flow and Metabolism, 8*(6), S4-S12.

[72] Nesterov, Y., Nemirovskii, A. (1994). In Society for Industrial and Applied Mathematics: Studies in Applied Mathematics, *Interior-Point Polynomial Algorithms in Convex Programming*. Philadelphia, PA.

[73] NVIDIA®. CUDA^{TM} GPUs. Retrieved 6/10, 2014, from https://developer.nvidia.com/cuda-gpus

[74] NVIDIA®. (2012). CUDA^{TM} Toolkit 4.2 CUSPARSE Library. Retrieved 6/10, 2014, from https://developer.download.nvidia.com/compute/DevZone/docs/html/CUDALibraries/doc/CUSPARSE_Library.pdf

[75] NVIDIA®. Compute Unified Device Architecture (CUDA^{TM}). Retrieved 6/10, 2014, from https://developer.nvidia.com/cuda-zone

[76] Nyquist, H. (1928). Certain topics in telegraph transmission theory. *Transactions of the American Institute of Electrical Engineers*. *47*(2)*,* 617-644.

[77] OpenMP. The OpenMP API specification for parallel programming. Retrieved 6/10, 2014, from http://www.openmp.org

[78] Parra, L. C., Barrett, H. H. (1998). List-mode likelihood: EM Algorithm and image quality estimation demonstration on 2-D PET. *IEEE Transactions in Medical Imaging, 17*, 228-235.

[79] Prekeges, J. (2013). Jones and Bartlett Learning, *Nuclear Instruments in Medicine (2nd Edition).* Burlington, MA: Jones and Bartlett.

[80] Pysz, M.A., Gambhir, S.S., Willmann, J.K. (2010). Molecular imaging: current status and emerging strategies. *Clinical Radiology, 65*(7), 500-516.

[81] Radon, Johann. (1917). Über die Bestimmung von funktionen durch ihre integralwerte laengs gewisser mannigfaltigkeiten. *Bergisch Sächsische Akademie der Wissenschaften 69*, 262-277.

[82] Rizzo, B. M., Haworth, S. T., Clough, A. V. (2014). Automated Lung Segmentation for Low Resolution CT Scans of Rats. *Proceedings of the SPIE: Medical Imaging 2014: Biomedical Applications in Molecular, Structural, and Functional Imaging*, San Diego, CA, USA., *9038*

[83] Rockafellar, T. R. (1976). Monotone operators and the proximal point algorithm. *SIAM Journal on Control and Optimization. 14*(5), 877-898.

[84] Rockafeller, T. R. (1970). *Convex Analysis*. Princeton, NJ: Princeton University Press.

[85] Saha, G. B. (2013). *Physics and Radiobiology of Nuclear Medicine*. New York, New York: Springer.

[86] Saha, G. B. (1992). Quality Control of Radiopharmaceuticals. *Fundamentals of nuclear pharmacy (3rd ed.), Springer NY, NY,* 143-167. doi: 10.1007/978-1-4757-4027-1_8

[87] Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal, 27*(3), 379-423.

[88] Shepp, L. A., Vardi, Y. (1982). Maximum likelihood reconstruction for emission tomography. *IEEE Transactions on Medical Imaging, 1*(2), 113-122. doi:10.1109/TMI.1982.4307558

[89] Shepp, L. A., Logan, B. F. (1974). The Fourier Reconstruction of a Head Section. *IEEE Transactions on Nuclear Science. 21*(3), 21-43.

[90][Si] Siddon, R. L. (1985). Fast calculation of the exact radiological path for a three−dimensional CT array. *Medical Physics, 12*(2), 252-255. doi: 0094-2405/85/020252-04$01.20

[91] Sidky, E. Y., Chartrand, R., Boone, J. M., & Pan, X. (2014). Constained TpV minimization for enhanced exploitation of gradient sparsity: Application to CT image reconstruction. *Medical Imaging and Diagnostic Radiology, 2*, 2168-2372. doi:10.1109/JTEHM.2014.2300862

[92] Sidky, E. Y., Jørgensen, J. H., & Pan, X. (2012). Convex optimization problem prototyping for image reconstruction in computed tomography with the Chambolle–Pock algorithm. *Physics in Medicine and Biology, 57*(10), 3065–3091. doi:10.1088/0031-9155/57/10/3065

[93] Sidky, E. Y., & Pan, X. (2008). Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization. *Physics in Medicine and Biology, 53*(17), 4777-4807. doi:10.1088/0031-9155/53/17/021

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

[95] Smith, M. F., & Jaszczak, R. J. (1998). An analytic model of pinhole aperture penetration for 3D pinhole SPECT image reconstruction. *Physics in Medicine and Biology, 43*(3), 761-775. doi:10.1088/0031-9155/43/4/006

[96] Strang, G. (2007). *Computational science and engineering*. Wellesley, MA: Wellesley-Cambridge Press.

[97] Suess E., Malessa S., Ungersbock K., Kitz P., Podreka I., Heimberger K., Hornykiewicz O., Deecke L. (1991). Technetium-99m-d,1-hexamethylpropyleneamine oxime (HMPAO) uptake and glutathione content in brain tumors. *Journal of Nuclear Medicine 32*(9), 1675-1681.

[98] Suga K., Uchisako H., Nishigauchi K., Shimizu K., Kume N., Yamada N., Nakanishi T. (1994). Technetium-99m-HMPAO as a marker of chemical and irradiation lung injury: experimental and clinical investigations. *Journal of Nuclear Medicine 35*(9), 1520-1527.

[99] Tao, T. (2008). Onsegear lecture. Retrieved from: http://www.ntnu.edu/onsager

[100] van der Have F., Vastenhouw B., Ramakers R.M., Branderhorst W., Krah J.O., Ji C., Staelens S.G., Beekman F.J. (2009). U-SPECT-II: An Ultra-High-Resolution Device for Molecular Small-Animal Imaging. *Journal of Nuclear Medicine, 50*(4), 599-605. doi: 10.2967/jnumed.108.056606

[101] Wheeler, A. P., Bernard, G. R. (2007). Acute lung injury and the acute respiratory distress syndrome: a clinical review. *The Lancet, 369*(9527), 1553-1564.

[102] Wietholt, C., Hsiao, I., & Chen, C. (2006). New ray-driven system matrix for small-animal pinhole-SPECT with detector blur, geometric response and edge penetration modeling. *IEEE Nuclear Science Symposium Conference Record,* San Diego, CA, USA. *, 6* 3414-3419. doi:10.1109/NSSMIC.2006.353735

[103] Wirgin, A. (2008). The Inverse Crime. *arXiv:math-ph/0401050*. Retrieved June 15, 2014, from the arXiv database.

[104] Wolf, P. A., Jørgensen, J. S., Schmidt, T. G., & Sidky, E. Y. (2013). Few-view single photon emission computed tomography (SPECT) reconstruction based on a blurred piecewise constant object model. *Physics in Medicine and Biology, 58*(16), 5629-5652. doi:0031-9155/13/000000+24

[105] Yakovlev, S. Silhouette mouse on white background. Retrieved 8/1, 2016, from https://www.123rf.com/photo_11006504_silhouette-mouse-on-white-background.html

[106] Zeng, G. L., Gullberg, G. T. (2000). Unmatched Projector/Backprojector Pairs in an Iterative Reconstruction Algorithm. *IEEE Transactions on Medical Imaging, 19*(5), 548-555.

[107] Zeng, G. L., Gullberg, G. T. (1991). Three-dimensional iterative reconstruction algorithms with attenuation and geometric point response correction. *IEEE Transactions on Nuclear Science, 38*(2), 693-702. doi:10.1109/23.289376