Fast Regularization Design for Tomographic Image Reconstruction for Uniform and Isotropic Spatial Resolution

by

Hugo R. Shi

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Electrical Engineering: Systems)
in The University of Michigan
2008

Doctoral Committee:

Professor Jeffrey A. Fessler, Chair
Professor Charles R. Meyer
Associate Professor Sandeep Pradhan Sandanandarao
Research Professor Neal H. Clinthorne

To my grandparents

ACKNOWLEDGEMENTS

First and foremost, I want to and must thank my advisor, Jeff Fessler. Jeff really is the best advisor a grad student could ask for. He is supportive, accommodating, and dedicated, even though he has so many other students, responsibilities, and a family. I think it is safe to say that without Jeff, not only would this dissertation not exist, but I might have been stuck in Ann Arbor for 1, 2, 5 or 10 years longer. Did I mention he is also a wild jazz musician?

I would also like to extend great thanks to my committee, Professor Neal H. Clinthorne, Professor Charles R. Meyer, and Professor Sandeep Pradhan Sandanandarao for their expertise and valuable feedback on this work. I would also like to thank all of my colleagues in, and out of Jeff's group. It has been a pleasure working with you all. I would also like to thank Professor Ron Fearing, Professor Jan Rabaey, Professor Kannan Ramchandran who have been a great help in my academic development.

I would also like to thank my family, my Grandparents, my Mom, my Dad, and my sister Karen for their love and support these past 5 years. Especially for bringing me home regularly to experience the joy that is California food. I would especially like to thank my Grandmother Lin, who truly inspires me to strive for great things. I would also like to thank Xuan Ying Wendy Tng for her love and support, and for taking care of me during the many tough times that make up grad school.

I would not have survived this experience if it were not for the Students of Color of Rackham, Thank you for providing the community that I needed to make it. I would also

like to thank the Detroit Asian Youth project, which has been one of the most rewarding experiences in my life, and has taught me a great deal about community organizing, youth, and education.

Last but not least, I would like to thank my friends from the Mood Makers Posse, who helped keep me sane, especially during those cold winter months. Especially Wendy, Jen and Cristina, who delivered food at my defense!.

Anyways, thank you all - truly and deeply, from the bottom of my heart.

TABLE OF CONTENTS

CHAPTER

LIST OF FIGURES

Figure

LIST OF APPENDICES

Appendix

LIST OF ABBREVIATONS

PET Positron Emission Tomography

CT Computed Tomography

FBP Filtered Back-projection

BPF Back-project Filter

ML Maximum-Likelihood

PL Penalized-Likelihood

WLS Weighted Least Squares

PWLS Penalized Weighted Least Squares

PULS Penalized Unweighted Least Squares

PSF Point Spread Function

LOR Line of Response

LIR Local Impulse Response

DFT Discrete Fourier Transform

AIMA Angular Integral Mostly Analytical

FIIN Full Integral Iterative NNLS

NNLS Non-Negative Least Squares

FWHM Full Width Half Max

RMS Root Mean Squared

HU Hounsfield Unit

KKT Karush-Kuhn-Tucker conditions

CHAPTER I

Introduction

Imaging systems are often evaluated based on the spatial resolution properties of reconstructed images. Systems with highly blurred spatial resolution can obscure features of interest to the viewer. Systems with extremely sharp spatial resolution give the viewer access to much detail, but may have too much noise. Many imaging systems are approximately shift-invariant; they exhibit the same spatial resolution properties in all regions of the resulting image. Images resulting from these systems can be thought of as the convolution between an impulse response and the true image. Physical aspects of positron emission tomography (PET) and computed tomography (CT) systems however, make them shift-variant. These imaging systems have spatial resolution properties that change through different parts of the resulting image. For example, fan-beam geometries for CT concentrate more rays at the edges of the image which result in sharper spatial resolution in that region. Crystal penetration effects in PET also introduce non-uniformities [29]. Spatial resolution variations can also result from the properties of the estimator used to reconstruct images. Estimators that use the shift-variant statistics of an image often have shift-variant spatial resolution properties.

Initial work on tomographic image reconstruction modeled tomographic systems as Radon transforms. This methodology assumes that data measurements consist of line integrals through the object. Using this system model, the reconstruction problem is reduced to inverting the Radon transform and can be solved by very quickly by filtered back projection (FBP). However actual tomographic systems contain intricacies that are not captured in the Radon transform. For example, the crystal penetration effects mentioned previously in PET systems can cause artifacts when regular FBP is used for reconstruction. Unfortunately taking these inaccuracies into consideration results in a reconstruction problem far more complicated than Radon transform inversion. Closed form solutions for such problems usually do not exist, or are very difficult to compute. There have been several attempts at analytical methods that compensate for space-variant physical effects [6, 30, 37, 62, 73, 75] but these generally ignore noise models. As a result, most reconstruction algorithms that use accurate system models are iterative. Iterative algorithms also have better spatial resolution [74] and noise properties [8] because they can model the measurement statistics. This can aid in lesion detection [48].

The spatial resolution of reconstructed images has a great impact on any tasks performed on those images. Anisotropic spatial resolution can distort the shape of features in the reconstructed image. This affects all tasks that involve the extraction or detection of specific shapes. Non-uniform resolution can distort the amplitude of reconstructed images in certain regions and complicate quantitation tasks. Some [47] argue that forcing isotropic resolution can hurt lesion detection and quantitation, though we believe the importance of uniform and isotropic spatial resolution requires further study. Furthermore, the techniques developed in this thesis can be used for other user defined spatial resolutions. We believe that the spatial resolution of the reconstructed image should be within the users control, and not an arbitrary byproduct of the estimator.

For iterative algorithms, Maximum likelihood (ML) reconstruction algorithms are often used for PET systems. In CT, due to the higher number of photon counts, weighted least squares (WLS) can be used, except perhaps at very low x-ray dose. Final converged solutions from ML and WLS reconstruction are often extremely noisy due to the ill conditioned nature of tomography. One can stop the iterative algorithm before convergence and before images become too noisy; however this solution does not achieve uniform and isotropic spatial resolution because resolution isotropy and uniformity can change with successive iterations [36, 72, 76]. Alternatively, one can run the iterative algorithm to completion and post-filter the resulting noise, however, this often requires a large number of iterations [61]. In this work we focus on using a roughness penalty in the iterative algorithm. Penalty functions lead to faster convergence of iterative algorithms. One can also choose penalty functions for different goals, including preserving edges [5, 27, 31, 34, 43], incorporating anatomical information [21, 33], and even lesion detection [49, 50]. Using penalty functions is advantageous because it allows us to predict the resolution and noise properties of the estimator and design the penalty accordingly [4,16,17,23,51,52,68,77]. Unfortunately, for penalized likelihood (PL) reconstruction, interactions between conventional regularizers (which are essentially just high-pass filters) and the Poisson log-likelihood function lead to nonuniform and anisotropic resolution in the resulting image [22, 23]. For Penalized WLS (PWLS), selecting weightings that match the implicit weightings associated with PL reconstruction result in good noise properties. In this case interactions between the weightings and conventional regularizers also lead to nonuniform and anisotropic resolution in the resulting image. The goal of this work is to design regularizers which reduce noise, lead to faster convergence, and preserve uniform and isotropic spatial resolution.

Much previous work on regularization design focuses on matrix-based approaches to fit the local impulse response of the estimator to a target impulse response. A shift-variant regularizer based on the aggregate certainty of measurement rays intersecting each pixel was developed that yielded uniform but anisotropic spatial resolution [23]. Stayman parameterized the quadratic regularizer to produce uniform and isotropic spatial resolution [65] and generalized regularization design to other non-Poisson noise models [67]. [40] presents a regularization design method for uniform and isotropic spatial resolution which is not based on an explicit target point spread function (PSF), but rather focuses on circular symmetry and uniformity. Fessler proposed an analytical approach to regularization design for 2D parallel-beam emission reconstruction that uses continuous space analogs to simplify the regularization design problem [18] which is the basis for much of this work. Qi and Leahy in [52] also proposed a regularization design method that provides a uniform peak amplitude of the local impulse response function for 3D PET, but that approach does not ensure isotropic resolution properties [52]. We have also proposed a solution for 3D PET that attempts to achieve uniformity and isotropy [58].

The main contributions of this work are:

- An analytical solution to regularization design problems for 2D parallel beam geometries is discussed in [18]. This work extends that analysis to fan-beam geometries by looking at a change of coordinate space for the weightings involved [59, 60];
- A computationally efficient approach [58] (but not a completely analytical solution) to 3D regularization design problems based on the work of [18, 65];
- A slower but more accurate regularization design approach that avoids the use of approximations in calculating the local frequency response of the regularizer;
- Several practical constraints on regularization design coefficients that improve the conditioning of the reconstruction problem and reduce artifacts in reconstructed images;
- A combined approach of 2D regularization and z-dimension post-reconstruction denoising to capture the resolution and computational efficient advantages of 2D regu-

larization for use in 3D geometries. We also characterize the spatial resolution and noise properties of this approach and investigate the performance of edge preserving regularization and denoising;

This thesis first describes the statistical model considered in Chapter §II. It then analyzes the local impulse response of the imaging system (which depends on our regularization design) in §2.5. We will formulate a frequency domain expression for the LIR of our iterative reconstruction using our regularizer for the 2D case in §3.2 and the 3D case in §3.3. Our target spatial resolution is also discussed in §3.2 and §3.3. Sections §4.1 and §4.2 discuss implementation details for computing regularization coefficients for the 2D and 3D case respectively. The results of these methods are presented in chapter §V. Chapter §VI presents reconstruction with 2D regularization and z-dimension iterative denoising which hopes to capture the isotropy and computational efficiency of 2D regularization and adapt it to 3D systems. Chapter §VII concludes the thesis and presents future work.

CHAPTER II

Background

2.1 X-ray Computed Tomography

X-ray systems consist of a source that emits 30-160keV x-ray photons that pass through an object being imaged. The photons either pass through the object and strike the detector, or interact with a part of the object and are absorbed or scatter (they may still strike the detector elsewhere, but this contributes to system noise). In the one dimensional case we consider an X-ray point source with all photons traveling in a line. The mean number of photons that travel a distance k without interacting with the object is:

$$E[N(k)] = N(0)e^{-\int_0^k \mu(k')dk'}.$$

Here, N(0) are the number of photons at x=0 (the number of total photons transmitted), and μ is the linear attenuation coefficient for the media that the x-ray photons travel through (usually air and the object being imaged). CT systems consist of an X-ray source (often modeled as a point source), and an array of detectors. The patient is placed between the X-ray source and the array of detectors, and then the system is rotated over some trajectory (usually circular or helical but other orbits have been investigated) to capture projection data. For CT systems, we use the following model for the measurements:

$$E[Y_i] = \int_\xi I_i(\xi)e^{-\int_{L_i} \mu(x,y,z,\xi)dl'}d\xi + r_i. \tag{2.1}$$

The subscript i denotes the ith measurement. ξ is the photon energy. The inner integral is no longer a function of depth k, but now a line integral corresponding to each measurement. The number of photons $N(0)$ has been replaced with $I_i(\xi)$, which is the intensity of the ith measurement incorporating the initial spectrum and the gain of the detector. r_i denotes background contributions coming from sources such as scatter. The line integral $\int_{L_i} \mu(x, y, z, \xi) dl'$ depends on the specific geometry of the imaging system but often approximates the Radon transform, a fundamental aspect of tomographic systems that will be discussed in §2.3.1.

Most CT systems use some sort of fan-beam geometry. An example of fan-beam geometry with a flat detector is pictured in Fig. 2.1. Fan beam geometries can induce resolution non-uniformities [28]. The spread of rays as they fan out from the source to the detector can magnify objects that are close to the detector. This work is focused on correcting non-uniformities and anisotropy caused by the statistical weightings used in reconstruction and the effects of the fan-beam geometry. This work can be extended to correct other shift-variant aspects of the system by using methods from [67].

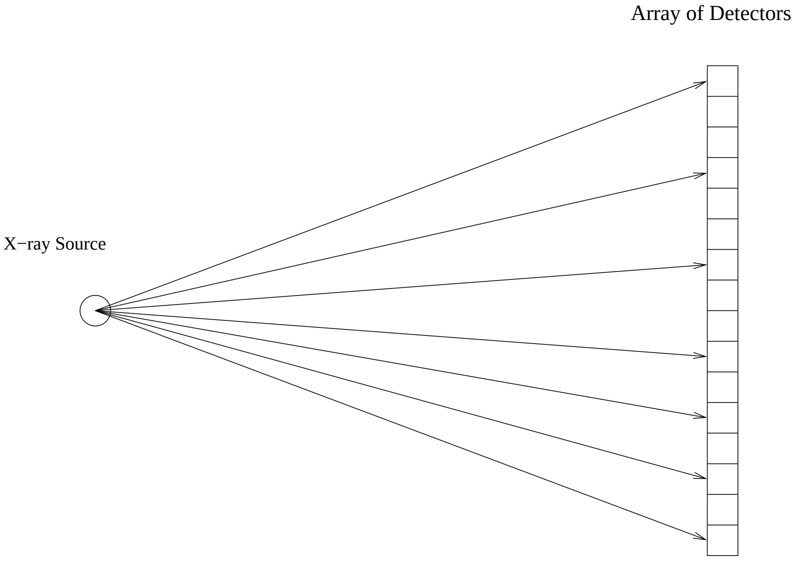


Figure 2.1: Diagram of Fan Beam 2D CT system.

2.2 Positron Emission Tomography

PET systems image the distribution of radioactive tracers in an object. In clinical situations, this is a human being with radioactive tracers injected inside of them. Positron decay creates 511keV photons that are emitted in anti-parallel directions. The patient is surrounded by a ring (2D systems) or a cylinder (3D systems) of detectors. Fig. 2.2 shows a diagram of a PET imaging system.

When a photon strikes a detector, the event is recorded as a detection. Simultaneous detections (or detections close enough in time to be considered simultaneous) are recorded as a coincidence. The number of coincidences between a pair of detectors is an indication of the amount of positron decay that occurred somewhere on the line between the 2 detectors. With some randomness, this is a measure of the line integral through that portion of the object. All line integrals along a certain angle are referred to as a projection. It should be noted that the raw sinogram data collected by a PET scanner indicates the specific location of radioactive tracers inside a patient during the scan duration. However, the specific location of radiotracers change over time and what is actually of interest is the distribution of radioactive tracers inside the patient. Inherent in every PET reconstruction problem is the estimation of the distribution of radio-tracers from the sinogram recorded during the time of the scan.

Photons used in PET have much higher photon energy compared to that of CT. Lower photon energies lead to better contrast between tissue types. This is important for CT because in CT, it is the attenuation of the tissues that is measured. However PET systems measure the distribution of radioactive tracers and the attenuation of other tissues is an obstacle to reconstruction. Mathematically, the ideal model for the mean of PET data we use is:

$$E[Y_i] = c_i \int_{L_i} f(x, y)dl + r_i.$$

$f(x, y)$ is the object of interest, and r_i are background events due to scatter and other effects. The line integral is taken for each line of response (LOR), which is the line between a pair of detectors. The specifics of the line integral are dependent on the physical geometries of the scanner, but like in CT, is usually based on the Radon transform §2.3.1 and the detector response. c_i accounts for the affects of attenuation in the body and represents the probability that the photons of a coincidence will not be blocked by the attenuation of the object. The probability that the coincidence will be detected is the product of probabilities of each photon surviving the attenuation of the object,

$$\begin{aligned} P_i &= P_{i,1} P_{i,2} \\ &= e^{\int_{L_{i,1}} \mu(x,y,z)dl} e^{\int_{L_{i,2}} \mu(x,y,z)dl}. \end{aligned}$$

Here, μ represents the attenuation of the body, which is a function of space, and the photon energy ξ. The two line integrals represent the paths of each photon. Since photons are emitted in anti-parallel directions, the two line integrals can be combined,

$$P_i = e^{\int_{L_i} \mu(x,y,z)dl},$$

which is just the line integral through the body. The attenuation can be measured accurately using a CT scan, which is part of the reason for PET-CT systems.

3D PET systems can be thought of as a stack of 2D systems that image multiple slices through an object. In 3D systems cross-plane coincidences, those that occur between detectors on different slices, can be detected. Some 3D PET systems include septa, which are physical barriers that block cross-plane coincidences, in order to reduce scatter. Using septa decreases the sensitivity of the imaging system due to the blocked coincidences,

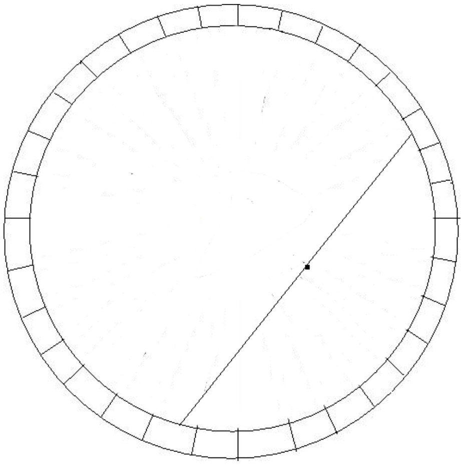


Figure 2.2: View of one slice of a 3D PET system

however septa also simplify the reconstruction process. Removing septa from PET systems increase their sensitivity substantially, e.g. by a factor of 7.04 [9]. This is caused by an increased number of lines of response (including cross-plane coincidences), the elimination of septa-shadowing effect (in some systems, septa also block some in-plane coincidences), and about a factor of 3 increase in scatter.

Various factors contribute to the shift-variant aspect of PET systems. Crystal penetration effects as well as the angles that photons strike detectors affect the aperture functions for PET detectors [56], resulting in non-uniform spatial resolution. The existence and geometry of septa can also cause changes in spatial resolution [26]. This work is designed to correct non-uniformities and anisotropy caused by the reconstruction process, not the system itself. To compensate for shift-invariant systems, one can apply the methods of [67].

2.3 Tomography

2.3.1 The Radon Transform

The Radon transform [53] is a set of line integrals through an object $f(x, y)$ set at angle ϕ and radius r from the center. Mathematically, it is expressed as:

$$g_{\phi}(r) = \int_x \int_y f(x, y)\delta(x\cos(\phi) + y\sin(\phi) - r)dxdy.$$

$g_\phi(\cdot)$ is referred to as a projection at angle ϕ. Data collected from parallel-beam CT systems and PET systems are naturally organized into projections. For parallel-beam CT systems, each angle where data is collected as the CT system rotates around the object being imaged creates a projection. For PET systems, LORs can be grouped by angle and thus form a projection. A set of projections is called a sinogram because Radon transforms of point objects look like sinusoids. A sample object and the sinogram of that object are shown in Fig. 2.3. The nature of the Radon transform leads to a simple way to reconstruct data from sinograms.

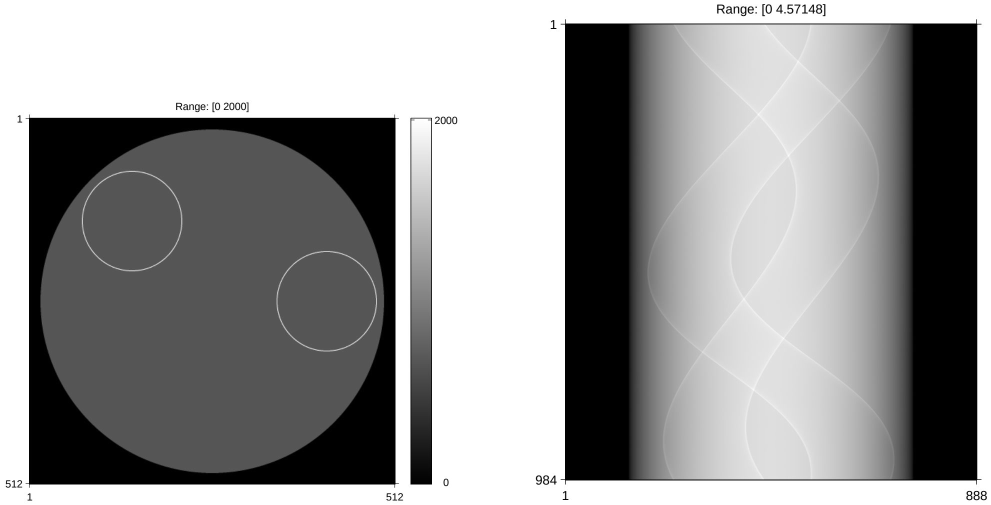


Figure 2.3: Left: A sample phantom., Right: sinogram of that phantom.

2.3.2 The Fourier Slice Theorem

The Fourier Slice Theorem states that 1D Fourier transform of $g_\phi(r)$ is the 2D Fourier transform of the object evaluated at angle ϕ. Mathematically,

$$G_\phi(\rho) = F(\rho, \Phi)|_{\Phi=\phi},$$

where Φ, ρ are frequency domain polar coordinates and G and F are Fourier transforms of $g_\phi(r)$ and $f(x, y)$ respectively. Consequently, we can also find $g_\phi(r)$ by taking the inverse Fourier transform of a 1D slice through F, evaluated at ϕ.

2.3.3 Backprojection-Filtering

One method of tomographic reconstruction is Backprojection-Filtering (BPF). Backprojection operators "smear" projection data back across the image. Mathematically,

$$b_\phi(x,y) = \int_{-\infty}^{\infty} g_\phi(r)\delta(x\cos\phi + y\sin\phi - r)dr. \tag{2.2}$$

The laminogram is the integral (or in real discrete systems, a sum) of all back-projections

$$f_b(x,y) = \int_0^{\pi} b_\phi(x,y)d\phi. \tag{2.3}$$

One can show that the laminogram is actually the image convolved with a $\frac{1}{r}$ blurring function: Plugging (2.2) into (2.3)

$$\begin{aligned} f_b(x,y) &= \int_0^{\pi} b_\phi(x,y)d\phi \\ &= \int_0^{\pi}\int_{-\infty}^{\infty} g_\phi(r)\delta(x\cos\phi + y\sin\phi - r)drd\phi. \end{aligned} \tag{2.4}$$

Using the Fourier slice theorem,

$$\begin{aligned} f_b(x,y) &= \int_0^{\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(\rho,\phi)e^{i2\pi\rho r}d\rho\delta(x\cos\phi + y\sin\phi - r)drd\phi \\ &= \int_0^{\pi}\int_{-\infty}^{\infty} F(\rho,\phi)e^{i2\pi\rho(x\cos\phi + y\sin\phi)}d\rho d\phi. \end{aligned} \tag{2.5}$$

This is nearly the inverse Fourier transform in polar coordinates,

$$f(x,y) = \int_0^{2\pi}\int_0^{\infty} \cdot\rho d\rho d\Phi$$

except that r should be integrated from $(0,\infty)$ and ϕ should be integrated from $(0,2\pi)$, and we need a ρ scale factor. In polar coordinates, $F(-\rho,\Phi) = F(\rho,\Phi+\pi)$, which means the two limits of integration are equivalent. Dividing by ρ:

$$f(x,y) = \int_0^{2\pi}\int_0^{\infty} \frac{F(\rho,\phi)}{\rho}e^{i2\pi\rho(x\cos\phi + y\sin\phi)}d\rho d\phi. \tag{2.6}$$

Thus, the laminogram is the original image blurred by a $1/\rho$ filter. Filtering the laminogram with a ρ filter can restore the original image. Real filters however, must be constructed with some frequency cutoff which will eliminate high frequency content in the image. In practice we often use Filtered Backprojection (FBP) instead of BPF. In FBP the order is reversed: filtering with a ramp filter first and then backprojecting the data. Though FBP reconstruction is very quick and efficient, FBP assumes that PET and CT systems behave like pure Radon transforms. Iterative reconstruction techniques allow for more accurate system models. Iterative techniques also allow one to use the noise statistics of the measurements in the estimation of the true image. Some have argued that statistically based post-filters could be used for FBP, however this is problematic because noise statistics are rooted in the sinogram domain. Converting these statistics to the image domain, which would be necessary for a post-filter, is a difficult task.

2.4 Iterative Algorithms

2.4.1 Statistical Models

For iterative reconstruction, we discretize the continuous functions in the previous sections as follows. For CT, let $\boldsymbol{y} = (y_1, ..., y_M)$ denote the vector of noisy sinogram measurements recorded in an emission or transmission tomography imaging system. For simplicity in the transmission tomography case, we consider the following mono-energetic formulation for the mean of our data:

$$\bar{y}_i = E[y_i] = b_i e^{-[\boldsymbol{A}\boldsymbol{x}]_i} + r_i. \tag{2.7}$$

For emission tomography we consider the model

$$\bar{y}_i = E[y_i] = c_i[\boldsymbol{A}\boldsymbol{x}]_i + r_i. \tag{2.8}$$

$\boldsymbol{A}$ is the system matrix, $\boldsymbol{x} = (x_1, x_2, ..., x_N)$ is a discretized version of the object being imaged, and $[\boldsymbol{A}\boldsymbol{x}]_i = \sum_{j=1}^{N} a_{ij} x_j$. In PET, The coefficients c_i denote attenuation coefficients. In CT, b_i denotes the blank scan. b_i is analogous to $I_i(\xi)$ in (2.1) and gives us the response of the system without any object attenuation present and quantifies the energy of the spectrum and the detector gain. r_i denotes random counts. In practice, c_i, b_i, and r_i are determined by various supplementary measurements. This work treats those quantities as predefined quantities and their estimation is not in the scope of this work. Note that all expressions thus far in this section (2.7), (2.8), are expressed as means. The measurements y_i are random variables, and it is this randomness that we refer to as noise, not the additive randoms r_i.

The radio-tracers in PET systems, and the X-ray source in CT are modeled as emitting photons randomly at a certain mean rate. Thus a Poisson model for the measurements makes sense. In CT, while the X-ray source may be Poisson, the distribution of the measurements is more complicated and a compound Poisson model is more accurate [12]. In this thesis we assume a Poisson distribution and a mono-energetic source for simplicity. In PET systems, positron decay is modeled as a Poisson process. All radio-tracers along a specific LOR are summed to a specific measurement. This measurement is a Poisson process as the sum of Poisson processes is also a Poisson process. Other noise models can be used and justified and are discussed in [67]. These statistical models are extremely simple and do not include many physical effects because the focus of this work is on correcting non-uniformities and anisotropy caused by statistical weightings, not the system itself.

2.4.2 Maximum and Penalized Likelihood Estimation

Once a noise model has been selected, one can use a Maximum-Likelihood (ML) estimator that picks the estimate which would maximize the probability of the measurements. Mathematically, this is expressed as.

$$\hat{\boldsymbol{x}}_{ML} = \arg\max_{\boldsymbol{x}} l(\boldsymbol{x}, \boldsymbol{y}).$$

Assuming a Poisson distribution, the likelihood of PET and CT measurements is

$$l(\boldsymbol{x}, \boldsymbol{y}) = \prod_i \frac{\bar{y}_i(\boldsymbol{x})^{y_i} e^{-\bar{y}_i(\boldsymbol{x})}}{y_i!}$$

Maximizing the log of the likelihood (or log-likelihood) is equivalent and often easier than maximizing the likelihood function itself. The corresponding log-likelihood function is

$$L(\boldsymbol{x}, \boldsymbol{y}) = \log l(\boldsymbol{x}, \boldsymbol{y}) = \sum_i y_i \log \bar{y}_i(\boldsymbol{x}) - \bar{y}_i(\boldsymbol{x}). \tag{2.9}$$

Unfortunately, reconstructing images based purely on the likelihood function results in extremely noisy images. As discussed in Chapter §I, there are several ways to reduce noise, but this work focuses on using regularization. Adding regularization we switch from ML reconstruction to Penalized Likelihood (PL) reconstruction where we maximize the objective function

$$\Phi(\boldsymbol{x}, \boldsymbol{y}) = L(\boldsymbol{x}, \boldsymbol{y}) - \zeta R(\boldsymbol{x}). \tag{2.10}$$

ζ is a user-selected factor that controls how much regularization affects the reconstructed image.

For CT, The log-likelihood $L(\boldsymbol{x}, \boldsymbol{y})$ is non-convex and non-quadratic, making the minimization problem difficult and slow. One can use a second-order Taylor-series approximation to the log-likelihood resulting in a PWLS cost function [13,14]. PWLS is ill-suited for PET reconstruction because PET involves much lower counts than CT, leading to a poor Taylor series approximation. In PWLS for CT, we estimate $\boldsymbol{x}$ by minimizing the following cost function:

$$\Phi(\boldsymbol{x}, \boldsymbol{y}) \triangleq ||\ell(\boldsymbol{y}) - \boldsymbol{A}\boldsymbol{x}||^2_{\boldsymbol{W}} + \zeta R(\boldsymbol{x}) \tag{2.11}$$

$$\hat{\boldsymbol{x}}(\boldsymbol{y}) = \arg\min_{\boldsymbol{x}} \Phi(\boldsymbol{x}, \boldsymbol{y}), \tag{2.12}$$

where $\ell(\boldsymbol{y}) \triangleq -\ln(\frac{\boldsymbol{y}-\boldsymbol{r}}{\boldsymbol{b}})$, $R(\boldsymbol{x})$ is a regularizer that controls noise, ζ is a scalar that determines the resolution-noise tradeoff, and $\boldsymbol{W} = \mathsf{diag}\{w_i\}$ is a weighting matrix.

The weighting matrix $\boldsymbol{W}$ is chosen to take advantage of the statistics of the data. In general, $L(\boldsymbol{y}, \boldsymbol{x}) = \sum h_i(y_i, [\boldsymbol{A}\boldsymbol{x}]_i)$ and properties of $h_i(\cdot, \cdot)$ determine $\boldsymbol{W}$ [67]. For transmission tomography, with the model in (2.7), $w_i = \mathsf{Var}^{-1}\{l_i\} \approx y_i$. For emission tomography, the plug-in weighting is $\boldsymbol{w}_i = \mathsf{Var}^{-1}\{y_i\} \approx y_i$. For low count levels alternative weightings are preferable [35]. The regularization design methods in this paper apply to other statistical models (such as compound Poisson) than (2.7) by simply changing the weighting matrix $\boldsymbol{W}$. In the next section §2.5, we will discuss local impulse responses (LIR) for PL and QPWLS. Both estimation methods result in similar expressions for the LIR when w_i is chosen as described here.

2.5 Local Impulse Response

For linear space-invariant systems, a global impulse response can be used to characterize the system. However with non-linear estimators we must analyze the local impulse response (LIR) to assess the uniformity and isotropy of the system. We use the following definition of the LIR, in a slightly different form than [23] at pixel or voxel j as:

$$l^j(\boldsymbol{x}) = \lim_{\epsilon \longrightarrow 0} \frac{\hat{\boldsymbol{x}}(\bar{\boldsymbol{y}}(\boldsymbol{x} + \epsilon \boldsymbol{\delta}^j)) - \hat{\boldsymbol{x}}(\bar{\boldsymbol{y}}(\boldsymbol{x}))}{\epsilon} \tag{2.13}$$

where $\boldsymbol{\delta}^j$ is an impulse function centered at pixel j. This is the limit of the difference between a reconstruction of the true image plus a small impulse perturbation, and a reconstruction of the true image without the perturbation, as the magnitude of the impulse approaches zero. Because we are trying to characterize the estimator, we use the mean of the data (a quantity that is not available in practice, but can be used in the analysis). This equation is equivalent to taking the derivative of the estimator which can be broken up using the chain rule:

$$l^j(\boldsymbol{x}) = \nabla \hat{\boldsymbol{x}}(\bar{\boldsymbol{y}}(\boldsymbol{x}))\boldsymbol{\delta}^j \tag{2.14}$$

$$= \nabla_{\boldsymbol{y}} \hat{\boldsymbol{x}}(\boldsymbol{y})\Big|_{\boldsymbol{y}=\bar{\boldsymbol{y}}(\boldsymbol{x})} \nabla_{\boldsymbol{x}} \bar{\boldsymbol{y}}(\boldsymbol{x})\Big|_{\boldsymbol{x}=\boldsymbol{x}} \boldsymbol{\delta}^j. \tag{2.15}$$

To calculate $\nabla_{\boldsymbol{y}} \hat{\boldsymbol{x}}(\boldsymbol{y})\Big|_{\boldsymbol{y}=\bar{\boldsymbol{y}}(\boldsymbol{x})}$ for the PL estimator, we can use the implicit function theorem and the chain rule [23]. We know that at the maximum of the objective function, the derivative of the objective function is zero,

$$\nabla^{[1,0]}\Phi(\hat{\boldsymbol{x}}(\boldsymbol{y}), \boldsymbol{y}) = 0,$$

where $\nabla^{[1,0]} = [\partial/\partial_{\hat{\boldsymbol{x}}_1}, ..., \partial/\partial_{\hat{\boldsymbol{x}}_p}]$ is the row gradient operator with respect to $\hat{\boldsymbol{x}}(\boldsymbol{y})$. If we differentiate again,

$$\nabla^{[2,0]}\Phi(\hat{\boldsymbol{x}}(\boldsymbol{y}), \boldsymbol{y})\nabla_{\boldsymbol{y}}\hat{\boldsymbol{x}}(\boldsymbol{y}) + \nabla^{[1,1]}\Phi(\hat{\boldsymbol{x}}(\boldsymbol{y}) = 0$$

$$\nabla_{\boldsymbol{y}}\hat{\boldsymbol{x}}(\boldsymbol{y}) = [-\nabla^{[2,0]}\Phi(\hat{\boldsymbol{x}}(\boldsymbol{y}), \boldsymbol{y})]^{-1}\nabla^{[1,1]}\Phi(\hat{\boldsymbol{x}}(\boldsymbol{y}). \tag{2.16}$$

Here, $\nabla^{[2,0]}$ is the Hessian operator whose (j,k)th element is $\frac{\partial}{\partial_{\hat{\boldsymbol{x}}_j}\partial_{\hat{\boldsymbol{x}}_k}}$, and $\nabla^{[1,1]}$ is the operator whose (j,i)th element is $\frac{\partial}{\partial_{\hat{\boldsymbol{x}}_j}\partial_{\boldsymbol{y}_i}}$. Plugging (2.10) into (2.16),

$$\nabla_{\boldsymbol{y}}\hat{\boldsymbol{x}}(\boldsymbol{y}) = [-\nabla^{[2,0]}L(\hat{\boldsymbol{x}}(\boldsymbol{y}), \boldsymbol{y}) + \boldsymbol{R}]^{-1}\nabla^{[1,1]}L(\hat{\boldsymbol{x}}(\boldsymbol{y}), \boldsymbol{y}), \tag{2.17}$$

where $\boldsymbol{R}$ is the Hessian of the regularizer $R(\boldsymbol{x})$. $\nabla^{[1,1]}R(\boldsymbol{x})$ is 0 since R is not a function of $\boldsymbol{y}$. Using a Poisson noise model [54] and (2.8) for emission tomography [15, 23],

$$-\nabla^{[2,0]}L(\hat{\boldsymbol{x}}(\boldsymbol{y}), \boldsymbol{y}) = \boldsymbol{A}' \,\mathsf{diag}\left\{\frac{y_i}{\bar{y}_i^2(\boldsymbol{x})}\right\}\boldsymbol{A}$$

$$-\nabla^{[1,1]}L(\hat{\boldsymbol{x}}(\boldsymbol{y}), \boldsymbol{y}) = \boldsymbol{A}' \,\mathsf{diag}\left\{\frac{1}{\bar{y}_i(\boldsymbol{x})}\right\}$$

$$\nabla_{\boldsymbol{x}}\bar{\boldsymbol{y}}(\boldsymbol{x})\Big|_{\boldsymbol{x}=\boldsymbol{x}}\boldsymbol{\delta}^j = \boldsymbol{A}\boldsymbol{\delta}^j. \tag{2.18}$$

In practical implementation, we approximate the mean of the data with a single realization, $\bar{y}_i(\boldsymbol{x}) \approx y_i$. Using this and combining (2.18), (2.17), and (2.15),

$$\boldsymbol{l}^j = \boldsymbol{l}^j(\boldsymbol{x}_{\text{true}}, \boldsymbol{R}) = [\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A} + \zeta\boldsymbol{R}]^{-1}\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A}\boldsymbol{\delta}^j, \tag{2.19}$$

where $\boldsymbol{R}$ is the Hessian of the regularizer $R(\boldsymbol{x})$, and $\boldsymbol{W} = \mathsf{diag}\{1/y_i\}$.

Using a Poisson noise model [55] and (2.7) for transmission tomography,

$$\begin{aligned} -\nabla^{[2,0]} L(\hat{\boldsymbol{x}}(\boldsymbol{y}), \boldsymbol{y}) &= \boldsymbol{A}' \,\mathsf{diag}\left\{ (\bar{y}_i(\boldsymbol{x}) - r_i)(1 - \frac{r_i y_i}{\bar{y}_i^2(\boldsymbol{x})}) \right\} \\ -\nabla^{[1,1]} L(\hat{\boldsymbol{x}}(\boldsymbol{y}), \boldsymbol{y}) &= \boldsymbol{A}' \,\mathsf{diag}\left\{ 1 - \frac{r_i}{\bar{y}_i(\boldsymbol{x})} \right\} \\ \nabla_{\boldsymbol{x}} \bar{\boldsymbol{y}}(\boldsymbol{x}) \Big|_{\boldsymbol{x}=\boldsymbol{x}} \boldsymbol{\delta}^j &= b_i e^{[\boldsymbol{A}\boldsymbol{x}]_i} \boldsymbol{A}\boldsymbol{\delta}^j = (\bar{y}_i(\boldsymbol{x}) - r_i)\boldsymbol{A}\boldsymbol{\delta}^j. \end{aligned} \tag{2.20}$$

Approximating the mean of the data with a single realization, $\bar{y}_i(\boldsymbol{x}) \approx y_i$, and combining (2.20), (2.17), and (2.15) results in (2.19) with $\boldsymbol{W} = \mathsf{diag}\left\{\frac{(y_i - r_i)^2}{y_i}\right\}$. If we ignore r_i, $\boldsymbol{W} = \mathsf{diag}\{y_i\}$.

As mentioned earlier, for CT we often use PWLS estimators. For the cost function (2.11), the PWLS estimator is linear in l and has the analytical solution $\hat{\boldsymbol{x}}(\boldsymbol{y}) = [\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A} + \zeta\boldsymbol{R}]^{-1}\boldsymbol{A}'\boldsymbol{W}\boldsymbol{\ell}$. This means that (2.13) can be evaluated for PWLS reconstruction to be $\boldsymbol{l}^j = [\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A} + \zeta\boldsymbol{R}]^{-1}\boldsymbol{A}'W\boldsymbol{A}\boldsymbol{\delta}^j$. In PWLS, one often chooses the weighting matrix $\boldsymbol{W}$ to be the inverse of the variance, $\mathsf{Var}^{-1}\{\boldsymbol{l}\} \approx y_i$ which is the same weighting matrix found when the LIR of PL reconstruction for CT is analyzed. Thus, PWLS and PL reconstruction for CT have similar resolution properties if $\boldsymbol{W}$ is defined appropriately. This allows us to use the same approach to regularization design for PL and PWLS estimators. Other statistical models are discussed in [67].

As is evident from (2.19), the local impulse response depends on the regularizer through its Hessian, $\boldsymbol{R}$. We would like to design $\boldsymbol{R}$ such that the local impulse response $\boldsymbol{l}^j$ matches a target response $\boldsymbol{l}^0$ at every pixel j. We could phrase this matrix optimization problem as:

$$\arg\min_{\boldsymbol{R}} \sum_j \left|\left| \boldsymbol{l}^j(\boldsymbol{x}_{\text{true}}, \boldsymbol{R}) - \boldsymbol{l}^0 \right|\right| . \tag{2.21}$$

This matrix formulation of the design problem seems intractable, so we will turn to the frequency domain to simplify the problem.

2.6 Regularization

Most of our work focuses on quadratic regularization because quadratic regularizers are approximately circulant and the resolution and variance can be analyzed and calculated easily [20]. Some have argued that quadratic regularizers perform as well as several non-quadratic regularizers for lesion detectability in emission tomography [45, 46].

Quadratic regularizers are of the form

$$R(\boldsymbol{x}) = \sum_{j=1}^{p} \sum_{k=1}^{p} w_{jk} \psi(x_j - x_k), \tag{2.22}$$

where j is a columnized pixel index and a function of n, m. For quadratic regularizers, $\psi(t) = \frac{t^2}{2}$. Quadratic regularizers of this form have a simple matrix form

$$R(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}'\boldsymbol{R}\boldsymbol{x} = \begin{cases} \sum_{l=1}^{p} \frac{1}{2}(w_{lj} + w_{jl}), & k = j \\ -w_{jk}, & k \neq j \end{cases}$$

In this thesis we assume that each voxel j has a local neighborhood whose differences are being penalized indexed by l. w_{jk} for pixel pairs outside the local neighborhood are 0. We use an alternate form for regularizers functionally equivalent but which has a better connection to Fourier analysis,

$$R(\boldsymbol{x}) = \sum_{n,m} \sum_{l=1}^{L} r_l^j \frac{1}{2} |c_l(n,m) ** x(n,m)|^2, \tag{2.23}$$

and

$$c_l(n,m) = \frac{1}{\sqrt{n_l^2 + m_l^2}} (\delta(n,m) - \delta(n - n_l, m - m_l)), \tag{2.24}$$

where n, m are 2d coordinates for images. For the 3d case

$$R(\boldsymbol{x}) = \sum_{n,m,z} \sum_{l=1}^{L} r_l^j \frac{1}{2} |c_l(n,m,z) *** x(n,m,z)|^2, \tag{2.25}$$

and

$$c_l(n,m,z) = \frac{1}{\sqrt{||\boldsymbol{n}_l||}} (\delta(n,m,z) - \delta(n - n_l, m - m_l, z - z_l)), \tag{2.26}$$

where $\boldsymbol{n}_l = (n_l, m_l, z_l)$. Regularizers of this form can be converted to the form in (2.22) by using $w_{jk} = \frac{r_l^j}{||\boldsymbol{n}_l||}$.

For CT scanners, quadratic regularization can be undesirable because it over-smoothes edges in the reconstructed images. Many forms of non-quadratic regularization have been proposed [5, 7, 24, 25, 27, 31, 34, 38, 43]. Non-quadratic regularizers use a different form of ψ for (2.22). A common penalty function for edge-preserving regularization is the Huber penalty, which is quadratic for values of $t < \triangle$ where $\triangle$ is some user defined threshold. Mathematically,

$$\psi(t) = \begin{cases} \frac{t^2}{2}, & |t| < \triangle \\ \triangle(|t| - \triangle) + \frac{\triangle^2}{2}, & |t| \geq \triangle \end{cases} .$$

2.7 Summary

This chapter briefly reviews PET and CT systems. It discusses the Radon transform, Fourier slice theorem, and the traditional way of reconstructing tomographic images, filtered back-projection. Next it moves on to the statistical model used in this paper and iterative reconstruction algorithms to obtain an expression for the LIR. Finally it discusses regularizers, both quadratic and non-quadratic edge preserving regularization. Next, we will discuss the Fourier analysis which is the basis of this work.

CHAPTER III

Frequency Domain Analysis of the Local Impulse Response

This section first reviews the use of discrete Fourier transforms for resolution analysis, leading to a computationally intensive approach to regularization design. We then consider continuous-space analogs that lead to simplified designs.

3.1 Discrete Fourier Analysis

In (2.19), the matrix $\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A}$ is approximately locally circulant near a pixel j of interest [10], and $\boldsymbol{R}$ can be designed to be locally circulant. Let $\boldsymbol{Q}$ denote an orthonormal discrete Fourier transform matrix centered at pixel j. Fourier analysis of local impulse responses can be complicated by complex exponentials caused by non-centered impulse functions. This re-centering eliminates the complex exponentials. $\boldsymbol{Q}$ is size $J \times J$, where J represents the total number of voxels in the image. For 2D systems,

$$\boldsymbol{Q}_{k,g} = \frac{1}{\sqrt{J}} e^{-2\pi i \left(\frac{1}{N}[\omega_{1,g} n_{k-j}] + \frac{1}{M}[\omega_{2,g} m_{k-j}]\right)}. \tag{3.1}$$

For 3D systems,

$$\boldsymbol{Q}_{k,g} = \frac{1}{\sqrt{J}} e^{-2\pi i \left(\frac{1}{N}[\omega_{1,g} n_{k-j}] + \frac{1}{M}[\omega_{2,g} m_{k-j}] + \frac{1}{Z}[\omega_{3,g} z_{k-j}]\right)}. \tag{3.2}$$

In these equations, w_1, w_2, and w_3 are spatial frequency components which are a function of a columnized frequency index g, n m and z are indexes into the image which are a function of the columnized pixel index j, and N, M and Z are the total number of pixels in each direction, and $J = NM$ for 2D systems and $J = NMZ$ for 3D systems. $n_{k-j} = n_k - n_j, m_{k-j} = m_k - m_j, z_{k-j} = z_k - z_j$, and $j = MNz_j + Nm_j + n_j, k = MNz_k + Nm_k + n_k$.

Then we can factor $\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A}\boldsymbol{\delta}_k \approx \boldsymbol{Q}'\boldsymbol{\Lambda}^j\boldsymbol{Q}\boldsymbol{\delta}_k$ for pixels k near j, where $\boldsymbol{\Lambda}^j = \mathsf{diag}\{\lambda_g^j\}$ and $\lambda^j \triangleq \sqrt{J}\boldsymbol{Q}\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A}\boldsymbol{\delta}^j$, and $\boldsymbol{R}\boldsymbol{\delta}_k \approx \boldsymbol{Q}'\boldsymbol{\Gamma}^j\boldsymbol{Q}\boldsymbol{\delta}_k$ for pixels k near j, where $\boldsymbol{\Gamma}^j = \mathsf{diag}\{\gamma_g^j\}$ and $\gamma^j \triangleq \sqrt{J}\boldsymbol{Q}\boldsymbol{R}\boldsymbol{\delta}^j$ [52], [10]. Then we can approximate the LIR in (2.19) as:

$$\boldsymbol{l}^j(\boldsymbol{x}) \approx [\boldsymbol{Q}'\boldsymbol{\Lambda}^j\boldsymbol{Q} + \zeta\boldsymbol{Q}'\boldsymbol{\Gamma}^j\boldsymbol{Q}]^{-1}\boldsymbol{Q}'\boldsymbol{\Lambda}^j\boldsymbol{Q}\boldsymbol{\delta}^j \tag{3.3}$$

$$= \boldsymbol{Q}'\left[\frac{\boldsymbol{\Lambda}^j}{\boldsymbol{\Lambda}^j + \zeta\boldsymbol{\Gamma}^j}\right]\boldsymbol{Q}\boldsymbol{\delta}^j, \tag{3.4}$$

where the matrices in the bracketed term are diagonal and the division operation is element-wise on the diagonal. These approximations are accurate only for row and column indices that are “sufficiently close” to voxel j. The terms in brackets in (3.4) correspond to the local frequency response of our estimator. We would like to match these to the frequency response $\boldsymbol{L}^0$ of a target PSF (which will be discussed in §3.2.1 and §3.3.1) as closely as possible, i.e., we want

$$\boldsymbol{L}^j \triangleq \frac{\boldsymbol{\Lambda}^j}{\boldsymbol{\Lambda}^j + \zeta\boldsymbol{\Gamma}^j} \approx \boldsymbol{L}^0. \tag{3.5}$$

Based on (3.5), one might consider a DFT formulation of the regularization design using the following minimization approach:

$$\hat{\boldsymbol{\Gamma}}^j = \operatorname*{arg\,min}_{\boldsymbol{\Gamma}\in\mathcal{T}} \left|\left|\boldsymbol{L}^0 - \frac{\boldsymbol{\Lambda}^j}{\boldsymbol{\Lambda}^j + \zeta\boldsymbol{\Gamma}^j}\right|\right|, \tag{3.6}$$

where $\boldsymbol{L}^0$ is the frequency response of $\boldsymbol{l}^0$ and $\mathcal{T}$ denotes the set of possible frequency responses for the regularizer limited by its structure that will be enumerated in §3.2.1 and

§3.3.1. Alternatively, as a preview to methods to come, we can cross multiply the terms in (3.5) yielding the simpler design criterion:

$$\hat{\mathbf{\Gamma}}^j = \operatorname*{arg\,min}_{\mathbf{\Gamma} \in \mathcal{T}} ||\boldsymbol{L}^0(\mathbf{\Lambda}^j + \zeta \mathbf{\Gamma}^j) - \mathbf{\Lambda}^j||, \tag{3.7}$$

which is similar to the formulation specified in [67]. Because $\mathbf{\Lambda}^j$ is the DFT of $\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A}\boldsymbol{\delta}^j$, calculating $\mathbf{\Lambda}^j$ requires one forward projection and backward projection per voxel. Thus, a regularization design based on (3.6) or (3.7) will be very slow.

3.2 2D Contiunous Space Analogs

Prior to [18], regularization design methods were based on discrete matrices. As shown in Appendix A, the continuous-space analog of $\mathbf{\Lambda}^j$ is the frequency response $\frac{w^j(\Phi)}{|\rho|}$, where $w^j(\Phi)$ is an expression that incorporates the Jacobian from the change of coordinates from parallel-beam to fan-beam geometry, and weights from $\boldsymbol{W}$ that correspond to rays that intersect pixel j at angle Φ. Substituting this into (3.5) and using continuous space analogs of $\mathbf{\Gamma}^j$ in frequency domain polar coordinates (ρ, Φ), where ρ represents radial frequency and Φ represents angular frequency yields the following expression for the continuous space analog of $\boldsymbol{L}^j$

$$\begin{aligned} \mathcal{L}^j(\rho, \Phi) &\approx \frac{w^j(\Phi)\frac{1}{|\rho|}}{w^j(\Phi)\frac{1}{|\rho|} + \zeta R^j(\rho, \Phi)} \\ &= \frac{w^j(\Phi)}{w^j(\Phi) + \zeta |\rho| R^j(\rho, \Phi)}. \end{aligned} \tag{3.8}$$

Here, $R^j(\rho, \Phi)$ is the local frequency response of the regularizer near pixel j. The continuous space analog of $\mathbf{\Lambda}^j$ simplifies (3.8) to provide a more efficient approach to regularization design than (3.7), as detailed in §4.1 and §4.2.

3.2.1 Regularization Structure

Regularizers control roughness by penalizing differences between neighboring pixels. Indexing the image $\boldsymbol{x}$ as a 2D function $x(n, m)$, we define a differencing function for a regularizer that penalizes the lth direction as

$$c_l(n,m) = \frac{1}{\sqrt{n_l^2+m_l^2}}(\delta(n,m) - \delta(n-n_l, m-m_l)), \tag{3.9}$$

where typically $(n_l, m_l) \in \{(0,1),(1,0),(1,1),(-1,1)\}$, corresponding to horizontal, vertical, and diagonal differences. A conventional quadratic regularizer can then be expressed as

$$R(\boldsymbol{x}) = \sum_{n,m}\sum_{l=1}^{L}\frac{1}{2}|c_l(n,m) ** x(n,m)|^2, \tag{3.10}$$

where $**$ denotes 2D convolution. This conventional regularizer assigns the same weight to the differences between each neighbor. For our proposed method, we make the regularizer spatially adaptive with the addition of weighting coefficients r_l^j as follows:

$$R(\boldsymbol{x}) = \sum_{n,m}\sum_{l=1}^{L} r_l^j \frac{1}{2}|c_l(n,m) ** x(n,m)|^2, \tag{3.11}$$

where j is the columnized pixel index which is a function of (n,m). The objective of this paper is to design coefficients $\{r_l^j\}$. To this end, we must analyze the local frequency response $R^j(\rho,\Phi)$ of the Hessian $\boldsymbol{R}$ of the space-variant regularizer (3.11).

Taking the Fourier transform of (3.9) yields:

$$\begin{aligned} |C_l(\omega_1,\omega_2)|^2 &= \frac{1}{n_l^2+m_l^2}|1-e^{-i(\omega_1 n_l + \omega_2 m_l)}|^2 \\ &= \frac{1}{n_l^2+m_l^2}(2-2\cos(\omega_1 n_l + \omega_2 m_l)). \end{aligned} \tag{3.12}$$

One can think about achieving isotropy intuitively in polar coordinates as eliminating angular dependence. Therefore we convert (3.12) to polar frequency coordinates to simplify the analysis. We use frequency and sampling relationships [41], $\omega_1 = 2\pi\Delta_x\rho\cos(\Phi)$ and $\omega_2 = 2\pi\Delta_y\rho\sin(\Phi)$, where Δ_x, Δ_y is pixel size, and (ρ,Φ) are polar frequency coordinates. For simplicity, we assume that $\Delta_x = \Delta_y = 1$. Then,

$$|C_l(\rho,\Phi)|^2 \approx \frac{1}{n_l^2+m_l^2}(2-2\cos(u_l(\rho,\Phi)))^2, \tag{3.13}$$

where

$$u_l(\rho,\Phi) = n_l 2\pi\Delta_x \rho\cos(\Phi) + m_l 2\pi\Delta_y \rho\sin(\Phi)\,. \tag{3.14}$$

Plugging (3.13) into (3.11) yields

$$R^j(\rho,\Phi) = \sum_{l=1}^{L} \frac{r_l^j}{n_l^2+m_l^2}(2-2\cos(u_l(\rho,\Phi)))$$

.

This is an accurate expression for the local frequency response of $\boldsymbol{R}$ which will be used to develop an accurate but slower regularization design method, the Full Integral Iterative NNLS, *FIIN*, method.

Going back to (3.12), we can use the following approximation, which we will refer to as the *AIMA* approximation,

$$2-2\cos(x) \approx x^2 \tag{3.15}$$

to develop a simpler expression for the local frequency response of $\boldsymbol{R}$ which will lead to the Angular Integral Mostly Analytical, *AIMA*, method. This approximation is worse for high frequencies, however regularization design using this approximation is viable because the approximation errors which are present in our equation for the local frequency response and target local frequency response cancel each other out. With this approximation, we have

$$\begin{aligned}|C_l(\omega_1,\omega_2)|^2 &= \frac{1}{n_l^2+m_l^2}(2-2\cos(\omega_1 n_l+\omega_2 m_l))\\ &\approx \frac{1}{n_l^2+m_l^2}(\omega_1 n_l+\omega_2 m_l)^2.\end{aligned} \tag{3.16}$$

Converting (3.16) to polar frequency coordinates as before.

$$\begin{aligned}|C_l(\omega_1,\omega_2)|^2 &\approx \frac{1}{n_l^2+m_l^2}(n_l 2\pi\Delta_x\rho\cos(\Phi)+m_l 2\pi\Delta_y\rho\sin(\Phi))^2\\ &= \frac{1}{n_l^2+m_l^2}(2\pi\rho)^2(n_l\cos(\Phi)+m_l\sin(\Phi))^2\\ &= (2\pi\rho)^2\cos^2(\Phi-\Phi_l),\end{aligned}$$

where $\Phi_l \triangleq \tan^{-1}\frac{m_l}{n_l}$. Combining this with (3.11), our final expression for the local frequency response of the regularizer near the jth pixel is

$$R^j(\rho,\Phi) = \sum_{l=1}^{L} r_l^j (2\pi\rho)^2\cos^2(\Phi-\Phi_l)\,. \tag{3.17}$$

For the usual choice of $L = 4$ and for (n_l, m_l) described below (3.9), we have $\Phi_l \in \{0, \pi/2, \pi/4, 3\pi/4\}$.

3.2.2 Target Local Frequency Response

Substituting (3.17) into (3.8) yields a simple expression for the local frequency response of a PWLS estimator. We want to design each regularization coefficient vector $\boldsymbol{r}^j = (r_1^j, ..., r_L^j)$ such that our frequency response matches that of the target as closely as possible. We select our target frequency response to be the local frequency response associated with a penalized unweighted least squares (PULS) estimator at the center of the field of view because we know the frequency response is isotropic there.

For an unweighted cost function and a parallel beam geometry, the continuous-space frequency response that is analogous to $\boldsymbol{QA'A\delta}^j$ is $\frac{1}{|\rho|}$. As shown in (A.17) in Appendix A, for uniform weights $(w_i = 1)$ we have the following local frequency response for fan-beam geometries

$$H^j(\rho,\Phi) = \frac{2}{J(s^j(\Phi))|\rho|}. \tag{3.18}$$

where $J(s^j(\Phi))$ is the Jacobian for the change of coordinates from parallel-beam to fan-beam geometries as defined in (A.12), and $s^j(\Phi)$ is an index into the sinogram based on pixel j and angle Φ. We set $s = 0$ in (A.17), which corresponds to the center pixel leading to the following target local frequency response:

$$\begin{aligned}\mathcal{L}^0(\rho) &\triangleq \frac{\frac{2}{J(0)|\rho|}}{\frac{2}{J(0)|\rho|} + \zeta R_0(\rho, \Phi)} \\ &= \frac{1}{1 + \zeta|\rho| 0.5 J(0) R_0(\rho, \Phi)},\end{aligned} \tag{3.19}$$

where $\mathcal{L}^0$ is the continuous space analog of $\boldsymbol{L}^0$ in (3.5). Without the *AIMA* approximation,

$$R_0(\rho, \Phi) = 4 - 2\cos(2\pi\rho\cos\Phi) - 2\sin(2\pi\rho\sin\Phi)\,. \tag{3.20}$$

Using the *AIMA* approximation,

$$R_0(\rho, \Phi) = (2\pi\rho)^2. \tag{3.21}$$

3.3 3D Continuous Space Analogs

§3.2 presented an expression for the continuous space analog of $\boldsymbol{\Lambda}^j$ for 2D systems. This analog still holds for 3D systems if we model them as a stack of 2D systems (which we can reasonably do for the purposes of regularization design if we assume small cross-plane angles in PET, or small cone angles in CT). This work is aimed at such limited 3D systems which we refer to as 2.5D, though regularizers designed using this method will work with varying degrees of efficacy for 3D systems depending on how large the cross-plane angles or cone angles are. More work will be needed to extend this work to fully 3D PET, helical, and cone-beam CT. Using 3D spherical polar coordinates, (ϱ, Φ, Θ) where ϱ is 3D radial frequency, Φ is in-plane angular frequency, and Θ is axial angular frequency, the continuous space analog of $\boldsymbol{\Lambda}^j$ is

$$\frac{w^j(\Phi)}{|\varrho_{xy}|}, \tag{3.22}$$

where $\varrho_{xy} = \varrho\cos(\Theta) = \rho$. Next we look at the structure of our regularizer to derive a continuous space analog of $\boldsymbol{\Gamma}^j$.

3.3.1 Regularization Structure

We consider a 3D roughness penalty with a form similar to the 2D regularizer from §3.2. In this section we will derive a minimization problem which can be solved for the desired regularizer. The differencing functions for our regularizer that penalizes the lth neighbor in 3D space is

$$c_l(n,n,z) = \frac{1}{||\boldsymbol{n}_l||}(\delta(n,m,z) - \delta(n-n_l, m-m_l, z-z_l))$$

where $\boldsymbol{n}_l = (n_l, m_l, z_l)$. Our conventional regularizer can be expressed as

$$R(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}'\boldsymbol{R}\boldsymbol{x} = \sum_{n,m,z}\sum_{l=1}^{L}\frac{1}{2}(c_l *** x(n,m,z))^2). \tag{3.23}$$

In this conventional regularizer, the difference between each neighbor receives the same penalty. We make the regularizer spatially adaptive with the addition of coefficients r_l^j. yielding:

$$R(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}'\boldsymbol{R}\boldsymbol{x} = \sum_{n,m,z}\sum_{l=1}^{L} r_l^j \frac{1}{2}(c_l *** x(n,m,z))^2), \tag{3.24}$$

where j is a columnized voxel index which is also a function of $\boldsymbol{n} = (n,m,z)$. Taking the Fourier transform of c_l,

$$\begin{aligned}|C_l(\omega_1,\omega_2,\omega_3)|^2 &= \frac{1}{||\boldsymbol{n}_l||^2}|1 - e^{-i(\omega_1 n_l + \omega_2 m_l + \omega_3 z_l)}|^2 \\ &= \frac{1}{||\boldsymbol{n}_l||^2}(2 - 2\cos(\omega_1 n_l + \omega_2 m_l + \omega_3 z_l)\,.\end{aligned}$$

Similar to the 2D derivation, we convert the above expression to cylindrical polar frequency coordinates using frequency sampling relationships [41],

$$\omega_1 = 2\pi\Delta_x\varrho\cos(\Phi)\cos(\Theta)$$

$$\omega_2 = 2\pi\Delta_y\varrho\sin(\Phi)\cos(\Theta)$$

$$\omega_3 = 2\pi\Delta_z\varrho\sin(\Theta)\,.$$

In 2D $\Delta_x = \Delta_y = 1$ for square pixels, which is also reasonable for 3D, however Δ_z is not necessarily 1. Then,

$$|C_l(\rho, \Phi, \Theta)|^2 \approx \frac{1}{||\boldsymbol{n}_l||}(2 - 2\cos(u_l(\rho, \Phi, \Theta)))^2, \tag{3.25}$$

where

$$u_l(\rho, \Phi, \Theta) = n_l 2\pi\Delta_x \varrho \cos(\Phi)\cos(\Theta) + m_l 2\pi\Delta_y \varrho \sin(\Phi)\cos(\Theta) + z_l 2\pi\Delta_z \varrho \sin(\Theta)\,. \tag{3.26}$$

Plugging (3.25) into (3.24) yields

$$R^j(\rho, \Phi, \Theta) = \sum_{l=1}^{L} \frac{r_l^j}{||\boldsymbol{n}_l||}(2 - 2\cos(u_l(\rho, \Phi, \Theta))).$$

This expression for the local frequency response of $\boldsymbol{R}$ will be used to derive the *FIIN* method in 3D. Using the *AIMA* approximation, we can derive a simpler expression for (3.25).

$$\begin{aligned}
|C_l(\omega_1, \omega_2, \omega_3)|^2 &= \frac{1}{||\boldsymbol{n}_l||^2}(2 - 2\cos(\omega_1 n_l + \omega_2 m_l + \omega_3 z_l) \\
&\approx \frac{1}{||\boldsymbol{n}_l||^2}(2 - 2(1 - \frac{1}{2}(\omega_1 n_l + \omega_2 m_l + \omega_3 z_l)^2) \\
&\approx \frac{1}{||\boldsymbol{n}_l||^2}(\omega_1 n_l + \omega_2 m_l + \omega_3 z_l)^2.
\end{aligned} \tag{3.27}$$

Converting to cylindrical polar frequency coordinates,

$$\begin{aligned}
|C_l(\omega_1, \omega_2, \omega_3)|^2 \approx \frac{1}{||\boldsymbol{n}_l||^2}&(n_l 2\pi\Delta_x \varrho \cos(\Phi)\cos(\Theta) \\
&+ m_l 2\pi\Delta_y \varrho \sin(\Phi)\cos(\Theta) + z_l 2\pi\Delta_z \varrho \sin(\Theta))^2 \\
= \frac{1}{||\boldsymbol{n}_l||^2}&(2\pi\varrho)^2(n_l\Delta_x \cos(\Phi)\cos(\Theta) \\
&+ m_l\Delta_y \sin(\Phi)\cos(\Theta) + z_l\Delta_z \sin(\Theta))^2.
\end{aligned}$$

Let $e(\Phi,\Theta) = (\cos\Theta\cos\Phi, \cos\Theta\sin\Phi, \sin\Theta)$ and $\boldsymbol{\Delta} = (\Delta_x, \Delta_y, \Delta_z)$, and then $|C_l(\rho,\Phi,\Theta)|^2$ simplifies to

$$\begin{aligned} |C_l(\rho,\Phi,\Theta)|^2 &= \frac{1}{||\boldsymbol{n}_l||^2}(2\pi\varrho)^2(||\boldsymbol{n}||e(\Phi,\Theta)\cdot[e(\Phi_l,\Theta_l)\odot\boldsymbol{\Delta}])^2 \\ &= (2\pi\varrho)^2(e(\Phi,\Theta)\cdot[e(\Phi_l,\Theta_l)\odot\boldsymbol{\Delta}])^2, \end{aligned} \tag{3.28}$$

where $\odot$ is element-wise multiplication. Substituting (3.28) into (3.24) results in the following continuous-space analog of $\boldsymbol{\Gamma}^j$:

$$R^j(\varrho,\Phi,\Theta) = (2\pi\varrho)^2\sum_{l=1}^{L} r_l^j(e(\Phi,\Theta)\cdot[e(\Phi_l,\Theta_l)\odot\boldsymbol{\Delta}])^2. \tag{3.29}$$

Using (3.22), (3.29), and (3.5) we can construct the following continuous-space analog of $\boldsymbol{L}^j$:

$$\begin{aligned} \mathcal{L}^j(\rho,\Theta,\Phi) &= \frac{\frac{w^j(\Phi)}{|\varrho_{xy}|}}{\frac{w^j(\Phi)}{|\varrho_{xy}|} + \zeta R(\varrho,\Phi,\Theta)} \\ &= \frac{w^j(\Phi)}{w^j(\Phi) + |\varrho_{xy}|\zeta R(\varrho,\Phi,\Theta)}. \end{aligned} \tag{3.30}$$

3.3.2 Target Local Frequency Response

We wish to design $R(\varrho,\Phi,\Theta)$ such that $\mathcal{L}^j$ matches some target $\mathcal{L}^0$ for all voxels j. This subsection describes an expression for $\mathcal{L}^0$. In §3.2 we considered the local frequency response associated with QPULS reconstruction that is isotropic at the center of the image for 2D systems. Unfortunately, this is not true for 3D systems. The purpose of this derivation is to find an expression that is isotropic, but with a mathematical structure that makes it an attainable target. The frequency response of QPULS is a good starting point.

The 2D continuous space analog to $\boldsymbol{QA'A\delta}^j$ for parallel beam geometries is $1/|\varrho_{xy}|$. This is an adequate approximation for the 3D continuous space analog to $\boldsymbol{QA'A\delta}^j$ once again assuming small cross plane angles for PET and small cone angles for CT. Thus, the frequency response of QPULS for 3D systems is

$$\mathcal{L}^0(\varrho, \Phi, \Theta) = \frac{\frac{1}{|\varrho_{xy}|}}{\frac{1}{|\varrho_{xy}|} + \zeta R_0(\rho, \Phi, \Theta)}$$

$$= \frac{1}{1 + \zeta |\varrho_{xy}| R_0(\rho, \Phi, \Theta)}. \tag{3.31}$$

The frequency response of a conventional regularizer (3.23) with $(n_l, m_l, z_l) \in \{(1,0,0), (0,1,0), (0,0,1)\}$, assuming $\Delta_x = \Delta_y = \Delta_z = 1$ (which is not a realistic assumption for actual scanners but is acceptable in the derivation of an isotropic target impulse response) without the *AIMA* approximation is

(3.32)

$$R_0(\rho, \Phi, \Theta) = 6 - 2\cos(2\pi\rho\cos\Phi\cos\Theta) - 2\cos(2\pi\rho\sin\Phi\cos\Theta) - 2\cos(2\pi\rho\sin\Theta)\,.$$

Using the *AIMA* approximation, this simplifies to

$$R_0(\rho, \Phi, \Theta) = (2\pi\varrho)^2 \tag{3.33}$$

Using the relationship $\varrho_{xy} = \varrho\cos\Theta$, we have:

$$\mathcal{L}^0(\varrho, \Theta) = \frac{1}{1 + \zeta |\varrho| \cos(\Theta)\, R_0(\rho, \Phi, \Theta)}, \tag{3.34}$$

which is clearly not isotropic because of the $\cos(\theta)$ in the denominator. In 3D, the local frequency response associated with uniform weighting and conventional regularization is not isotropic. We use a similar structure and eliminate the anisotropic terms to create an isotropic 3D target,

$$\mathcal{L}^0(\varrho) = \frac{1}{1 + \zeta |\varrho| (2\pi\varrho)^2}. \tag{3.35}$$

The $\cos(\Theta)$ term in (3.34) represents the fact that for a 2.5D geometry, $\boldsymbol{A'WA\delta}^j$ is completely confined in the XY plane. Conventional quadratic regularizers that apply the same penalty in all directions preserve that anisotropy. In 3D systems, spatially-variant regularization design is even more important to compensate for the naturally ellipsoidal spatial resolution that these systems produce. In the next chapter, we will try to match the local frequency response with the target local frequency response using the following type of optimization approach:

$$\boldsymbol{r}^j = \arg\min_{\boldsymbol{r} \geq 0} ||\mathcal{L}^j - \mathcal{L}^0||^2.$$

CHAPTER IV

Solving for Regularization Coefficients

4.1 2D Problem

We try to design regularization coefficients $\boldsymbol{r}^j = (r_1^j, ..., r_L^j)$ to match our designed local frequency response (3.8) to the target local frequency response (3.19) as follows:

(4.1)

$$\boldsymbol{r}^j = \arg\min_{\boldsymbol{r} \geq 0} \int \int \left(\frac{w^j(\Phi)}{w^j(\Phi) + \zeta|\rho| R^j(\rho, \Phi)} - \frac{1}{1 + \zeta|\rho| 0.5 J(0) R_0(\rho, \Phi)} \right)^2 \mathcal{W}(\rho, \Phi) d\rho d\Phi,$$

where $\mathcal{W}(\rho, \Phi)$ is a weighting function. To derive a simpler minimization problem, we solve for $\boldsymbol{r}^j$ by trying to match the local frequency responses as follows:

$$\frac{w^j(\Phi)}{w^j(\Phi) + \zeta|\rho| R^j(\rho, \Phi)} \approx \frac{1}{1 + \zeta|\rho| 0.5 J(0) R_0(\rho, \Phi)}. \tag{4.2}$$

Next we will go over the derivation of the *FIIN* and *AIMA* method.

4.1.1 *FIIN* Method

This section derives the Full Integral Iterative NNLS regularization design method which integrates over all variables in the frequency domain and must be solved using an iterative NNLS algorithm. Cross multiplying (4.2) yields

$$w^j(\Phi) + w^j(\Phi)\zeta|\rho| 0.5 J(0) R_0(\rho, \Phi) \approx w^j(\Phi) + \zeta|\rho| R^j(\rho, \Phi)$$

$$w^j(\Phi) 0.5 J(0) R_0(\rho, \Phi) \approx R^j(\rho, \Phi).$$

Using the inner product $\langle f,\, g\rangle = \frac{1}{\pi}\int_0^\pi f(\Phi)g(\Phi)d\Phi$, we project $\tilde{w}^j(\Phi)R_0(\rho,\Phi))$ onto the space spanned by $\{1-\cos(u_l(\rho,\Phi))\}$ which can be orthonormalized into L basis functions, $\{p_l\}$ using Gram-Schmidt. Then our regularization design problem simplifies to

$$\boldsymbol{r}^j = \underset{\boldsymbol{r}\geq 0}{\arg\min}\, ||\boldsymbol{T}\boldsymbol{r} - \boldsymbol{b}^j||^2, \tag{4.3}$$

where $\boldsymbol{b}^j$ is a vector of inner products between $\tilde{w}^j(\Phi)R_0(\rho,\Phi))$ and the L orthonormal basis functions $\{p_l\}$, $b^j_k = \int\int p_k(\rho,\Phi)\tilde{w}^j(\Phi)R_0(\rho,\Phi)d\rho d\Phi$, and $\boldsymbol{T}$ is a $L \times L$ matrix whose elements are the inner products between $2 - 2\cos(u_l(\rho,\Phi))$ and $\{p_1, p_2, ...p_L\}$, $\boldsymbol{T}_{mn} = \int\int(2 - 2\cos(u_m(\rho,\Phi)))p_n(\rho,\Phi)d\rho d\Phi$. We define $\boldsymbol{P}$ to be an operator whose columns are $\{p_1, p_2, ...p_L\}$ then, $R^j(\rho,\Phi) = \boldsymbol{P}\boldsymbol{T}\boldsymbol{r}^j$. This design problem is then solved with an NNLS algorithm. This more accurate version is slower than the *AIMA* method that will be presented next because there is no apparent analytical solution similar to the one presented in §4.1.4.

4.1.2 *AIMA* Method

This section derives the Angular Integral Mostly Analytical, *AIMA*, method. This method uses the *AIMA* approximation to eliminate dependence on ρ and requires integrals over the angular variable of the frequency domain. This simplification leads to a mostly analytical solution which is computationally efficient. Starting from (4.2) and using the *AIMA* approximation (3.15),

$$w^j(\Phi)\zeta|\rho|0.5J(0)(2\pi\rho)^2 \approx \zeta|\rho|R(\rho,\Phi)$$

$$w^j(\Phi)0.5J(0)(2\pi\rho)^2 \approx \sum_{l=1}^{L} r^j_l (2\pi\rho)^2\cos^2(\Phi - \Phi_l)$$

$$\tilde{w}^j(\Phi) \triangleq w^j(\Phi)0.5J(0) \approx \sum_{l=1}^{L} r^j_l \cos^2(\Phi - \Phi_l)\,. \tag{4.4}$$

We design $\boldsymbol{r}^j$ by minimizing the difference between both sides of (4.4):

$$\boldsymbol{r}^j = \underset{\boldsymbol{r} \geq 0}{\arg\min} \frac{1}{\pi} \int_0^{\pi} (\tilde{w}^j(\Phi) - \sum_{l=1}^{L} r_l \cos^2(\Phi - \Phi_l))^2 d\Phi. \tag{4.5}$$

Minimizing (4.5) does not guarantee that we minimize (4.1) under an unweighted norm. However (4.1) and (4.5) are equivalent if we use a very specific weighted norm:

$$\mathcal{W}(\rho, \Phi) = ([w^j(\Phi) + \zeta|\rho| R(\rho, \Phi)][1 + \zeta|\rho| 0.5 J(0)(2\pi\rho)^2])^{\frac{1}{2}}. \tag{4.6}$$

Because of the high-pass nature of regularizers represented by the ρ^2 terms in the equation, this is a weighting that emphasizes high frequencies. This makes sense for this minimization problem first, because impulse responses are high frequency in nature, and second because at DC, when $\rho = 0$, the two sides of (4.2) match exactly. It should also be noted that (4.5) is independent of ρ, because terms based on ρ are identical in both local frequency responses and are eliminated in the simplification. This is consistent with an intuitive understanding of regularization design. We are trying to design LIRs that are isotropic, or independent of direction, so our efforts should be concentrated on eliminating angular dependence.

We solve for L coefficients using the above minimization (4.5) for each pixel j. We constrain the L coefficients to be non-negative, which is an easy way to ensure that the penalty function is convex, though other approaches have been considered [66]. This expression also applies to parallel-beam geometries, where $J(0) = 1$ and we remove the 0.5 scale factor because we integrate Φ over $[0, \pi)$ instead of $[0, 2\pi)$ as in fan-beam geometries.

We can think of the minimization in (4.5) as a projection of $\tilde{w}^j(\Phi)$ onto the space spanned by $\{\cos^2(\cdot - \Phi_l)\}$, which allows us to greatly simplify the problem and derive a computationally efficient analytical solution to the regularization design problem. Expanding these cosines to a three term basis that is orthonormal with respect to the inner product $\langle f,\, g\rangle = \frac{1}{\pi}\int_0^\pi f(\Phi)g(\Phi)d\Phi$ yields

$$\cos^2(\Phi - \Phi_l) \;=\; \frac{1}{2}p_1(\Phi) + \frac{\cos(2\Phi_l)}{2\sqrt{2}}p_2(\Phi) + \frac{\sin(2\Phi_l)}{2\sqrt{2}}p_3(\Phi), \tag{4.7}$$

where the three orthonormal basis functions are

$$\begin{aligned} p_1(\Phi) &= 1 \\ p_2(\Phi) &= \sqrt{2}\cos(2\Phi) \\ p_3(\Phi) &= \sqrt{2}\sin(2\Phi)\,. \end{aligned}$$

Using (4.5), we write $\sum_{l=1}^{L} r_l^j \cos^2(\Phi - \Phi_l) = \boldsymbol{PTr}^j$, where $\boldsymbol{P}$ is a matrix whose columns are p_1, p_2, and p_3, and $\boldsymbol{T}$ is a $3 \times L$ matrix of linear combination coefficients whose lth column is $[\frac{1}{2} \;\; \frac{\cos(2\Phi_l)}{2\sqrt{2}} \;\; \frac{\sin(2\Phi_l)}{2\sqrt{2}}]^T$. $\boldsymbol{T}$ is computed by taking the dot products of $\cos^2(\Phi - \Phi_l)$ and p_1, p_2, and p_3, $[T]_{mn} = \int \cos^2(\Phi - \Phi_m)\, p_n(\Phi)d\Phi$.

Using (4.7), the minimization problem (4.5) simplifies to the following expression:

$$\boldsymbol{r}^j = \operatorname*{arg\,min}_{\boldsymbol{r}\geq 0} ||\boldsymbol{Tr} - \boldsymbol{b}||^2, \tag{4.8}$$

where $\boldsymbol{P}^*$ denotes the adjoint of $\boldsymbol{P}$ and $\boldsymbol{b} \triangleq \boldsymbol{P}^*\tilde{\boldsymbol{w}}^j(\cdot)$, i.e., $b_k = \int p_k(\Phi)\tilde{w}^j(\Phi)d\Phi$, $k = 1, 2, 3$.

4.1.3 Zeroes in the Hessian

If there are too many zeros in $\boldsymbol{r}^j$, there will be zeros in the Hessian, possibly degrading $\hat{\boldsymbol{x}}$. This can cause elongated impulse responses that may contribute to streak artifacts in the reconstructed image. This phenomenon occurred in [18] however we did not notice the artifacts due to the coarser spatial resolution in PET. It is present when using *AIMA* with the ring phantom in §5.1. The problem improves when using *FIIN* method, thus we believe this phenomenon is a caused by the *AIMA* approximation as well as the non-negativity constraint. For *AIMA*, and to be safe using *FIIN* we modify (4.8) to ensure that an adequate number of r_l^j values are greater than some $\epsilon_l^j > 0$. Requiring the penalty coefficients for the vertical and horizontal directions be non-zero is sufficient to eliminate zeros in the Hessian (a similar constraint could be created using the 2 diagonals instead of the vertical and horizontal neighbors, however the *AIMA* approximation is worse for diagonal neighbors). We turn to previous work to select ϵ_l^j. In [23], we derived a certainty based weighting using a spatially variant $\kappa^j = \sqrt{\frac{\sum_i a_{ij}^2 \tilde{w}_i}{\sum_i a_{ij}^2}}$ that seeks to provide uniform spatial resolution. In terms of the continuous space analogs used in this paper, $\kappa^j = \sqrt{\frac{1}{\pi}\int_0^\pi \tilde{w}^j(\Phi)d\Phi}$. This regularization design method can be implemented using the regularization structure presented in this paper by setting $r_l^j = (\kappa^j)^2$ for $(m_l, n_l) = (0,1),(1,0)$. This approach provides a convenient nominal value for the regularization strength at each pixel. We define the lower constraint vector $\boldsymbol{\epsilon}^j$ such that $\epsilon_l^j = \alpha(\kappa^j)^2$ for vertical and horizontal neighbors, $(m_l, n_l) = (0,1),(1,0)$, and $\epsilon_l^j = 0$ for all other neighbors. Using a nonzero coefficient α can be thought of as controlling the tradeoff between certainty based regularization and the proposed *AIMA* and *FIIN* methods. For the results presented in this paper, we used $\alpha = 0.1$.

Now we formulate our problem so that non-negative least squares (NNLS) algorithms will accommodate this new constraint. Let $\bar{\boldsymbol{r}}^j \triangleq \boldsymbol{r}^j + \boldsymbol{\epsilon}^j$. Solving with the constraint of $\boldsymbol{r}^j \geq 0$ ensures that $\bar{\boldsymbol{r}}^j \geq \boldsymbol{\epsilon}^j$. Substituting $\bar{\boldsymbol{r}}^j$ into (4.8) yields $\boldsymbol{T}\bar{\boldsymbol{r}}^j - \boldsymbol{b}^j = \boldsymbol{T}(\boldsymbol{r}^j + \boldsymbol{\epsilon}^j) - \boldsymbol{b}^j = \boldsymbol{T}\boldsymbol{r}^j - (\boldsymbol{b}^j - \boldsymbol{T}\boldsymbol{\epsilon}^j)$. So our final cost function for regularization design is

$$\begin{aligned} \hat{\boldsymbol{r}}^j &= \underset{\boldsymbol{r}\geq 0}{\arg\min}\, ||\boldsymbol{T}\boldsymbol{r} - (\boldsymbol{b}^j - \boldsymbol{T}\boldsymbol{\epsilon}^j)|| \\ &= \underset{\boldsymbol{r}\geq 0}{\arg\min}\, ||\boldsymbol{T}\boldsymbol{r} - \bar{\boldsymbol{b}}^j||, \end{aligned} \tag{4.9}$$

where $\bar{\boldsymbol{b}}^j \triangleq \boldsymbol{b}^j - \boldsymbol{T}\boldsymbol{\epsilon}^j$. We use coefficients $\bar{\boldsymbol{r}}^j = \hat{\boldsymbol{r}}^j + \boldsymbol{\epsilon}^j$ in the regularizer (3.11). We next solve (4.9) analytically. It is this analytical solution that makes the fast regularization design technique more efficient than the *FIIN* method.

4.1.4 Analytical Solution

Using a second-order neighborhood $(L = 4)$ we select (n_l, m_l) to be $(1,0), (0,1), (1,1), (1,-1)$ leading to the following Φ_l: $\Phi_1 = 0,\ \Phi_2 = \pi/2,\ \Phi_3 = \pi/4,\ \Phi_4 = -\pi/4$. So the terms in (4.9) are

$$\begin{gathered}
\boldsymbol{T} = \tfrac{1}{2}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1/\sqrt{2} & -1/\sqrt{2} & 0 & 0 \\ 0 & 0 & 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix} \\
\bar{\boldsymbol{b}} = \begin{bmatrix} d_1 \\ \sqrt{2}d_2 \\ \sqrt{2}d_3 \end{bmatrix}, \ \boldsymbol{d} = \begin{bmatrix} \frac{1}{\pi}(1-\alpha)\int_0^{\pi} w^j(\Phi) d\Phi \\ \frac{1}{\pi}\int_0^{\pi} w^j(\Phi)\cos(2\Phi)\, d\Phi \\ \frac{1}{\pi}\int_0^{\pi} w^j(\Phi)\sin(2\Phi)\, d\Phi \end{bmatrix}.
\end{gathered} \tag{4.10}$$

Observe that $\boldsymbol{T}\boldsymbol{\epsilon}^j = [\alpha(\kappa^j)^2\ 0\ 0]^T$, so the effect of $\boldsymbol{\epsilon}^j$ is entirely contained to d_1. Ignoring α, d_1 is the continuous space analog of $(\kappa^j)^2$, d_2 is related to the horizontal and vertical directions, and d_3 is related to the diagonal directions.

This is an under-determined system, which is somewhat intuitive since one can obtain approximately isotropic smoothing using only the horizontal and vertical neighbors, or only the diagonal neighbors. For the purposes of regularization design, an under-determined situation is desirable since it allows us to use the "extra" degrees of freedom to ensure non-negativity even when anisotropic regularization is needed.

We could solve the minimization (4.9) using an iterative *NNLS algorithm* [32, p. 158]. However, using the properties of $\boldsymbol{T}$ and $\boldsymbol{d}$, we can avoid iterations almost entirely by solving (4.9) *analytically* using the KKT conditions. When $\alpha = 0$, $\sqrt{d_2^2 + d_3^2} \leq d_1$. as outlined in appendix §B. This inequality is usually true for the values of α used (typically around 0.1), however for pixels where it is not, the problem would have to be solved using

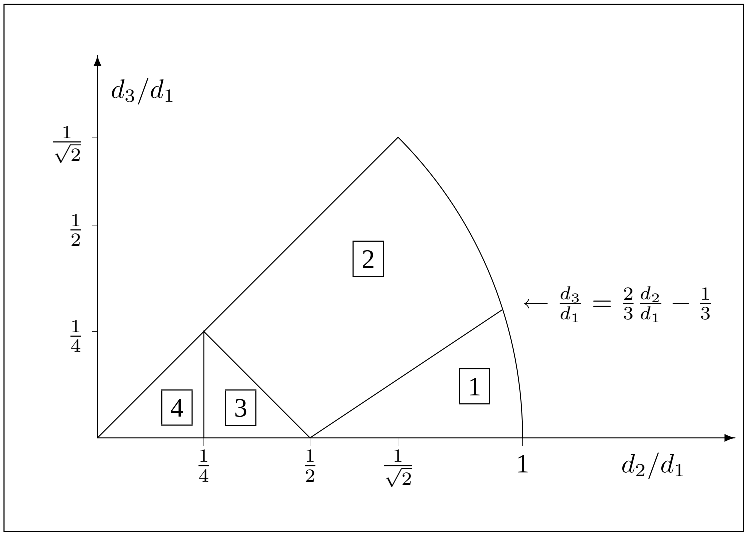


Figure 4.1: First octant of quadratic penalty design space showing the four regions where different constraints are active.

a NNLS algorithm rather than this analytical solution. For the phantoms, data, and values of α used in this thesis, this inequality was always true and we could use the analytical solution. For more details on using NNLS to solve the regularization problem, see §4.2.3.

The structure of $\boldsymbol{T}$ leads to eight-fold symmetry that simplifies analysis. If $d_2 < 0$ we can solve for $\boldsymbol{r}$ using $|d_2|$ and then swap r_1 with r_2. If $d_3 < 0$ we can solve for $\boldsymbol{r}$ using $|d_3|$ and then swap r_3 with r_4. If $d_3 > d_2$ we can solve for $\boldsymbol{r}$ with d_2 and d_3 interchanged, and then swap r_1 with r_3 and r_2 with r_4. Therefore, hereafter we focus on cases where $0 \leq d_3 \leq d_2 \leq d_1$. Fig. 4.1 shows these first octant cases, numbered according to the number of nonzero elements of $\boldsymbol{r}$.

$\boxed{1}$ If $d_2 \geq \frac{1}{2}d_1$ and $d_3 \leq \frac{2}{3}d_2 - \frac{1}{3}d_1$, then

$$r_1 = \frac{4}{3}\left(d_1 + d_2\right), \qquad r_2 = r_3 = r_4 = 0.$$

$\boxed{2}$ If $d_3 \geq \frac{2}{3}d_2 - \frac{1}{3}d_1$ and $d_3 + d_2 \geq \frac{1}{2}d_1$, then

$$r_1 = \frac{8}{5}\left[\frac{1}{2}d_1 + \frac{3}{2}d_2 - d_3\right], \quad r_2 = r_4 = 0,$$

$$r_3 = \frac{12}{5}\left[d_3 - \left(\frac{2}{3}d_2 - \frac{1}{3}d_1\right)\right].$$

$\boxed{3}$ If $d_3 + d_2 \leq \frac{1}{2}d_1$ and $d_2 \geq \frac{1}{4}d_1$, then

$r_1 = \frac{8}{5}\left[\frac{1}{2}d_1 + \frac{3}{2}d_2 - d_3\right], \quad r_2 = r_4 = 0,$

$r_3 = \frac{12}{5}\left[d_3 - \left(\frac{2}{3}d_2 - \frac{1}{3}d_1\right)\right].$

$\boxed{3}$ If $d_3 + d_2 \le \frac{1}{2}d_1$ and $d_2 \ge \frac{1}{4}d_1$, then there are multiple non-negative $\boldsymbol{r}$ that exactly solve $\nabla\,\Psi(\boldsymbol{r}) = \mathbf{0}$. The minimum-norm solution is $r_1 = 4d_2, \quad r_2 = 0,$
$r_3 = d_1 - 2d_2 + 2d_3, \quad r_4 = 2\left[\frac{1}{2}d_1 - (d_2 + d_3)\right].$

$\boxed{4}$ If $d_2 \le \frac{1}{4}d_1$, then there are multiple non-negative $\boldsymbol{r}$ that are exact solutions. The natural choice is the minimum-norm $\boldsymbol{r}$ given by the pseudo-inverse solution $\boldsymbol{r} = \boldsymbol{T}^{\dagger}\boldsymbol{d}$, where $r_1 = 2\left(\frac{1}{4}d_1 + d_2\right), \quad r_2 = 2\left(\frac{1}{4}d_1 - d_2\right),$
$r_3 = 2\left(\frac{1}{4}d_1 + d_3\right), \quad r_4 = 2\left(\frac{1}{4}d_1 - d_3\right).$

The analytical solution presented above is for the usual first-order differences. For higher-order differences or neighborhoods, it would appear to become increasingly cumbersome to solve (4.5) analytically, so an iterative NNLS approach may be more appealing. This can still be practical since $\boldsymbol{T}$ is quite small. The analytical solution above is a continuous function of $\boldsymbol{d}$, which in turn is a continuous function of $\tilde{w}^j(\Phi)$. This continuity property would seem to be desirable for avoiding artifacts in the reconstructed images. Without it, our $\boldsymbol{r}^j$ might not be smooth, thus violating our circulant approximation.

For practical implementation, we simply discretize the integrals in (4.10) [19]. This presents interpolation issues in extracting $\tilde{w}^j(\Phi[n_a])$, a discretized version of $\tilde{w}^j(\Phi)$, from $\boldsymbol{W}$ for $n_a = 1, ..., N_a$. For the parallel beam case, $\tilde{w}^j(\Phi[n_a])$ can be computed back-projections. The system matrix $\boldsymbol{A}$ is composed of blocks,

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_1 \\ \boldsymbol{A}_2 \\ \cdots \\ \boldsymbol{A}_{N_a} \end{bmatrix}, \tag{4.11}$$

where N_a is the number of angles used by the system and $\boldsymbol{A}_{n_a}\boldsymbol{x}$ computes one projection at angle $\phi[n_a]$. The back-projection $\boldsymbol{A}'$, which is the adjoint of $\boldsymbol{A}$ is also composed of blocks

$$\boldsymbol{A}' = [\boldsymbol{A}_1' \boldsymbol{A}_2' ... \boldsymbol{A}_{N_A}]. \tag{4.12}$$

The following expression generates weightings for angle n_a over all voxels,

$$\tilde{w}(\phi[n_a]) = \frac{[\mathbf{0}...\boldsymbol{A}_{n_a}...\mathbf{0}]\boldsymbol{y}}{[\mathbf{0}...\boldsymbol{A}_{n_a}...\mathbf{0}]\mathbf{1}}, \tag{4.13}$$

where the division is element-wise division and the denominator is a normalization term to control for scale factors in $\boldsymbol{A}$. For the fan beam case, there is an analytical formula for $\tilde{w}^j(\Phi[n_a])$ from $\boldsymbol{W}$ presented in Appendix §A. Those equations are presented as continuous space coordinates s which spans the length of the detector and ζ, the angle of rotation of the system. In discrete implementation, we round these off to the nearest neighbor to get indices into $\boldsymbol{W}$. The analytical solution presented in this section is very efficient and the bulk of the computation time is spent computing $\tilde{w}^j(\Phi)$ which has the compute time of approximately one back-projection. For further speed optimizations, down-sampled back-projections can be explored. In practice, $\tilde{w}^j(\Phi)$ is fairly smooth and the basis functions that we use to approximate $\tilde{w}^j(\Phi)$ are even smoother. One can achieve good results by calculating $\tilde{w}^j(\Phi[n_a])$ for fewer angles and interpolating the rest. The performance and efficacy of such a scheme still needs to be explored.

4.2 3D Problem

4.2.1 *FIIN*

As in the 2D problem, we match the expression for the LIR of the 3D system to the target LIR yielding

$$\begin{aligned} \frac{w^j(\Phi)}{w^j(\Phi) + \zeta|\varrho| \cos(\Theta)\, R^j(\rho, \Phi, \Theta)} &\approx \frac{1}{1 + \zeta(2\pi)^2 |\varrho|^3} \\ \zeta|\varrho| \cos(\Theta)\, R^j(\rho, \Phi, \Theta) &\approx \zeta w^j(\Phi)(2\pi)^2 |\varrho|^3 \\ \cos(\Theta)\, R^j(\rho, \Phi, \Theta)) &\approx w^j(\Phi)(2\pi|\varrho|)^2 \end{aligned}$$

(4.14)

Here, $\cos(\Theta)\, R^j(\rho, \Phi, \Theta)$ can be decomposed into L basis functions that can be orthonormalized using Gramm-Schmidt. The design problem can be solved in a way similar to §4.1.1.

4.2.2 *AIMA*

Starting from (4.14),

$$\begin{aligned} \cos(\Theta)\, R^j(\rho, \Phi, \Theta)) &\approx w^j(\Phi)(2\pi|\varrho|)^2 \\ (2\pi\varrho)^2 \sum_{l=1} r_l^j \cos\Theta(e(\Phi, \Theta) \cdot [e(\Phi_l, \Theta_l) \circ \boldsymbol{\Delta}])^2 &\approx w^j(\Phi)(2\pi|\varrho|)^2 \\ \sum_{l=1} r_l^j \cos\Theta(e(\Phi, \Theta) \cdot [e(\Phi_l, \Theta_l) \circ \boldsymbol{\Delta}])^2 &\approx w^j(\Phi) \end{aligned} \tag{4.15}$$

where $\circ$ represents element wise multiplication. To solve for the penalty coefficients r_l^j, we solve the following minimization problem

$$\boldsymbol{r} = \operatorname*{arg\,min}_{\boldsymbol{r} \geq 0} \int_0^{\pi} \int_{-\pi/2}^{\pi/2} |w^j(\Phi) - \sum_{l=1} r_l^j \cos\Theta(e(\Phi, \Theta) \cdot [e(\Phi_l, \Theta_l) \circ \boldsymbol{\Delta}])^2|^2 d\Theta, d\Phi. \tag{4.16}$$

This minimization problem involves projecting onto the space spanned by $\cos\Theta(e(\Phi, \Theta) \cdot [e(\Phi_l, \Theta_l) \circ \boldsymbol{\Delta}])^2$. We can decompose this term $\sum_{l=1} r_l^j \cos\Theta(e(\Phi, \Theta) \cdot e(\Phi, [e(\Phi_l, \Theta_l) \circ \boldsymbol{\Delta}])^2$ as $\boldsymbol{P}\boldsymbol{T}\boldsymbol{r}^j$ where r is $L \times 1$ vector of penalty coefficients. The term $\cos\Theta(e(\Phi, \Theta) \cdot [e(\Phi_l, \Theta_l) \circ \boldsymbol{\Delta}])^2$ can be expanded into 6 orthonormal basis functions which will make up the columns of $\boldsymbol{P}$. $\boldsymbol{T}$ is a matrix of linear combinations coefficients such that $\sum_{l=1} r_l^j \cos\Theta(e(\Phi, \Theta) \cdot$

$[e(\Phi_l,\Theta_l)\circ\mathbf{\Delta}])^2 = \boldsymbol{PTr}^j$. We first expand $[e(\Phi_l,\Theta_l)\circ\mathbf{\Delta}])^2$ into 6 orthogonal basis functions. After multiplying $(e(\Phi,\Theta)\cdot[e(\Phi_l,\Theta_l)\circ\mathbf{\Delta}])^2$ by $\cos\Theta$, all the functions are still orthogonal except for 2 which can be orthogonalized by Gram-Schmidt using the inner product $\langle f,\, g\rangle = \frac{1}{\pi^2}\int_{-\pi/2}^{\pi/2}\int_0^{\pi} f(\Phi,\Theta)g(\Phi,\Theta)d\Phi d\Theta$. Then we normalize to make them orthonormal. Hereafter we assume $\Delta_x = \Delta_y = 1$ (square but not cube voxels) which is realistic for scanners. Starting with $(e(\Phi,\Theta)\cdot[e(\Phi_l,\Theta_l)\circ\mathbf{\Delta}])^2$:

$$\begin{aligned}
(e(\Phi,\Theta)\cdot[e(\Phi_l,\Theta_l)\circ\mathbf{\Delta}])^2 &= \cos^2(\Theta_l)\cos^2(\Phi_l)\cos^2(\Theta)\cos^2(\Phi)\\
&+ \cos^2(\Theta_l)\sin^2(\Phi_l)\cos^2(\Theta)\sin^2(\Phi)\\
&+ \sin^2(\Theta_l)\,\Delta_z^2\sin^2(\Theta)\\
&+ 2\cos^2(\Theta_l)\cos(\Phi_l)\sin(\Phi_l)\cos^2(\Theta)\cos(\Phi)\sin(\Phi)\\
&+ 2\cos(\Phi_l)\cos(\Theta_l)\sin(\Theta_l)\,\Delta_z\cos(\Phi)\cos(\Theta)\sin(\Theta)\\
&+ 2\sin(\Phi_l)\cos(\Theta_l)\sin(\Theta_l)\,\Delta_z\sin(\Phi)\cos(\Theta)\sin(\Theta)\,.
\end{aligned}$$

There are 6 terms in this equation, corresponding to 6 basis functions, however they are not orthonormal. After making these equivalences using trigonometric identities:

$$\begin{aligned}
\sin^2(\Theta) &= \frac{1}{2}-\frac{1}{2}\cos(2\Theta)\\
\cos^2(\Theta)\cos^2(\Phi) &= \frac{1}{4}[1+\cos(2\Theta)+\cos(2\Phi)+\cos(2\Theta)\cos(2\Phi)]\\
\cos^2(\Theta)\sin^2(\Phi) &= \frac{1}{4}[1+\cos(2\Theta)-\cos(2\Phi)-\cos(2\Theta)\cos(2\Phi)],
\end{aligned}$$

we can extract 6 orthogonal basis functions

$$\begin{aligned}
p_1' &= 1 \\
p_2' &= \sin(\Phi)\sin(\Theta)\cos(\Theta) \\
p_3' &= \cos(\Phi)\sin(\Theta)\cos(\Theta) \\
p_4' &= \cos(2\Theta) \\
p_5' &= \cos(2\Phi) + \cos(2\Phi)\cos(2\Theta) \\
p_6' &= \cos^2(\Theta)\cos(\Phi)\sin(\Phi)\,.
\end{aligned}$$

After multiplying by $\cos\Theta$, we get

$$\begin{aligned}
p_1' &= \cos(\Theta) \\
p_2' &= \sin(\Phi)\sin(\Theta)\cos^2(\Theta) \\
p_3' &= \cos(\Phi)\sin(\Theta)\cos^2(\Theta) \\
p_4' &= \cos(2\Theta)\cos(\Theta) \\
p_5' &= [\cos(2\Phi) + \cos(2\Phi)\cos(2\Theta)]\cos(\Theta) \\
p_6' &= \cos^3(\Theta)\cos(\Phi)\sin(\Phi)\,.
\end{aligned}$$

Unfortunately, p_1' and p_4' are not orthogonal anymore, so we use Gram-Schmidt, yielding

$$\begin{aligned}
p_1' &= \cos(\Theta) \\
p_4' &= \cos(2\Theta)\cos(\Theta) - \langle \cos(2\Theta)\cos(\Theta), \frac{p_1'}{||p_q'||}\rangle \frac{p_1'}{||p_1'||}.
\end{aligned}$$

Finally we normalize all basis functions, leaving us with $\{p_1, p_2, ..., p_6\}$

We can now solve for $\boldsymbol{T}$ such that $\sum_{l=1} r_l \cos\Theta(e(\Phi,\Theta)\cdot[e(\Phi_l,\Theta_l)\circ\boldsymbol{\Delta}])^2 = \boldsymbol{PTr}^j$. Starting with the terms from the expansions in (4.17),

$$\begin{aligned}
t'_1 &= \cos^2(\Theta_l)\cos^2(\Phi_l) \\
t'_2 &= \cos^2(\Theta_l)\sin^2(\Phi_l) \\
t'_3 &= \sin^2(\Theta_l)\,\Delta_z^2 \\
t'_4 &= 2\cos^2(\Theta_l)\cos(\Phi_l)\sin(\Phi_l) \\
t'_5 &= 2\cos(\Phi_l)\cos(\Theta_l)\sin(\Theta_l)\,\Delta_z \\
t'_6 &= 2\sin(\Phi_l)\cos(\Theta_l)\sin(\Theta_l)\,\Delta_z.
\end{aligned}$$

Letting $k = \frac{1}{||1/2\cos\Theta\cos 2\Theta||}$, the lth column of $\boldsymbol{T}$ is

$$\begin{bmatrix}
(0.25t'_1(1+k) + 0.25t'_2(1+k) + 0.5't_3(1-k))1/||p'_1|| \\
(t'_6)1/||p'_2|| \\
(t'_5)1/||p'_3|| \\
(0.25t'_1 + 0.25t'_2 - 0.5t'_3)1/||p'_4|| \\
(0.25(t'_1 - t'_2))1/||p'_5|| \\
(t'_4)1/||p'_6||.
\end{bmatrix}$$

Using (4.17) and (4.17), the minimization problem (4.16) simplifies to the following expression:

$$\boldsymbol{r}^j = \underset{\boldsymbol{r}\geq 0}{\arg\min}\, ||\boldsymbol{T}\boldsymbol{r} - \boldsymbol{b}||^2, \tag{4.17}$$

where $\boldsymbol{P}^*$ denotes the adjoint of $\boldsymbol{P}$ and $\boldsymbol{b} \triangleq \boldsymbol{P}^*\boldsymbol{w}^j(\cdot)$, i.e., $b_k = \int p_k(\Phi)w^j(\Phi)d\Phi$, $k = 1, 2, 3, 4, 5, 6$.

4.2.3 Implementation Details

The minimization problems in (4.5) and (4.16) are under-determined and have the potential to yield many different r_l^j that are local minima. For our assumption of locally circulant matrices to be accurate in (3.4), the mapping from $w_j(\Phi)$ to $\boldsymbol{r}^j$ must be a continuous function. The analytical solution presented in §4.1 satisfies this constraint, however those methods have not been extended to the 3D case because the KKT conditions are more complicated in the 3D so we solve them by some NNLS algorithm. We can use Tikhonov regularization [71] to coerce NNLS to choose a solution that also minimizes the norm of $\boldsymbol{r}^j$.

We modify (4.16) by appending a scaled identity matrix to the bottom of $\boldsymbol{T}$ and zero padding $\boldsymbol{b}$

$$\tilde{\boldsymbol{T}} = \begin{pmatrix} \boldsymbol{T} \\ \tau \boldsymbol{I} \end{pmatrix}, \tilde{\boldsymbol{b}} = \begin{pmatrix} \boldsymbol{b} \\ \boldsymbol{0} \end{pmatrix},$$

so that our new minimization problem is

$$\underset{\boldsymbol{r}^j \geq 0}{\arg\min} \, ||\tilde{\boldsymbol{b}} - \tilde{\boldsymbol{T}}\boldsymbol{r}^j||. \tag{4.18}$$

τ is a very small number (in our implementation we used 0.001) that is sufficient to apply a slight penalty to the norm of $\boldsymbol{r}^j$. As in the 2D case, we should use the modifications in §4.1.3 to avoid having too many zeroes in the Hessian. In the 3D case we define the lower constraint vector $\boldsymbol{\epsilon}^j$ such that $\epsilon_l^j = \alpha(\kappa^j)^2$ for 3 adjacent neighbors, $(m_l, n_l, z_l) = (0,1,0), (1,0,0), (0,0,1)$, and $\epsilon_l^j = 0$ for all other neighbors. Here, $\kappa^j = \sqrt{\frac{\sum_i a_{ij}^2 \tilde{w}_i}{\sum_i a_{ij}^2}}$. The results of these methods are presented in the next chapter.

CHAPTER V

Results

In this chapter, we provide and discuss the results of various simulations using 2D and 3D regularization.

5.1 2D results

We first investigated imaging a phantom consisting of two uniform rings that highlight the effects of non-uniformities and anisotropy [40]. Afterwards, we studied real CT data.

5.1.1 Ring Phantom

We simulated a 2D 3rd-generation fan-beam CT system using distance-driven forward and backprojections [11]. The rotation center is 40.8cm from the detector, and the source is 94.9cm from the detector. The axis of rotation is at the center of the object. The simulated imaging system has 888 rays per view spaced 1mm apart, and 984 evenly spaced view angles over a full 2π rotation. The reconstructed images consisted of a 512×512 grid of 1 mm pixels. We chose a ζ in (3.19) such that the target PSF has a full width half max (FWHM) of 3.18 mm.

We simulated a noiseless fan-beam sinogram without scatter using a phantom consisting of a background disk and two rings each of thickness 1mm shown in Fig. 5.1. We generated the sinogram by taking the appropriate line integrals using the same system geometry through the analytical phantom. Fig. 5.2 shows penalty coefficients $\{\boldsymbol{r}^j\}$ designed using the *AIMA* method that yielded the following results. The top images show coefficients in the horizontal and vertical neighbors, and the bottom images show coefficients for diagonal neighbors. There is substantial spatial variation in these coefficients.

We reconstructed images $\hat{\boldsymbol{x}}$ using a conjugate-gradient algorithm with 40 iterations. using several regularizers. We selected a target PSF to be that associated with PULS at the center of the field of view. We first created an image uniformly blurred by the target PSF to serve as our target for this study. We then reconstructed images using, (i) conventional regularized weighted least squares reconstruction with $\boldsymbol{r}^j = 1$, $(m_l, n_l) = (0,1), (1,0)$, (ii) regularization using the certainty based weighting from [23], (iii) the *AIMA* method with $\alpha = 0.1$, (iv) the *AIMA* method with $\alpha = 0$, and (vi) the *FIIN* method with $\alpha = 0$. Fig. 5.3 is a closeup of the right-most ring reconstructed with the various methods listed above and Fig. 5.4 is a closeup of the left-most ring.

Fig. 5.5 and Fig. 5.6 show profiles around the two rings of the reconstructed images using the various regularization methods relative to the mean intensity of the rings from our target, PULS reconstruction with conventional regularization. This verifies that *AIMA* and *FIIN* improve resolution uniformity. Here, 0 radians corresponds to the rightmost point of that ring and the angles are measured clockwise.

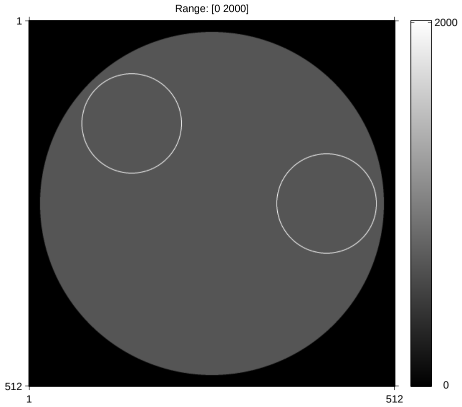


Figure 5.1: Ring phantom used for reconstruction.

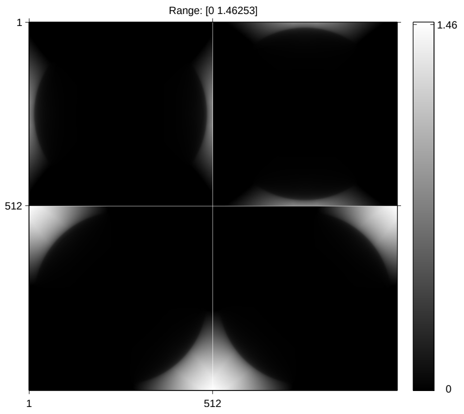


Figure 5.2: Regularization penalty coefficients used in reconstruction of ring phantom. The four images are r_l^j for $l = 1, ..., 4$.

5.1.2 Real CT Data

The second study used a similar imaging geometry using one slice of real CT data from a GE scanner described in [70] with weightings $\boldsymbol{W}$ computed in the same way as [69]. ζ was selected such that the target PSF had a FWHM of 1.51 mm.

Fig. 5.7 displays an image created with PWLS using conventional regularization with the impulse response locations denoted with crosses and a region marked that we will zoom in on. Fig. 5.8 - 5.12 show windowed reconstructions using conventional regularization, certainty based regularization, *AIMA* with $\alpha = 0.1$ and $\alpha = 0$, and *FIIN* with $\alpha = 0$ that have been zoomed in to the region marked in Fig. 5.7.

Fig. 5.13-5.16 show local impulse responses for the five regularization methods at several locations calculated analytically using (2.19). These figures show from left to right, the target impulse response, and local impulses responses for conventional regularization, certainty based regularization [23], *AIMA* with $\alpha = 0.1$, *AIMA* with $\alpha = 0$, and *FIIN* with $\alpha = 0$. Contour plots of the LIR are displayed below at 0.9, 0.75, 0.5, 0.25, and 0.1 of the maximum value of the target PSF. The LIR becomes more anisotropic near the edge of the FOV. Our Fourier-based regularization scheme compensates for this anisotropy better

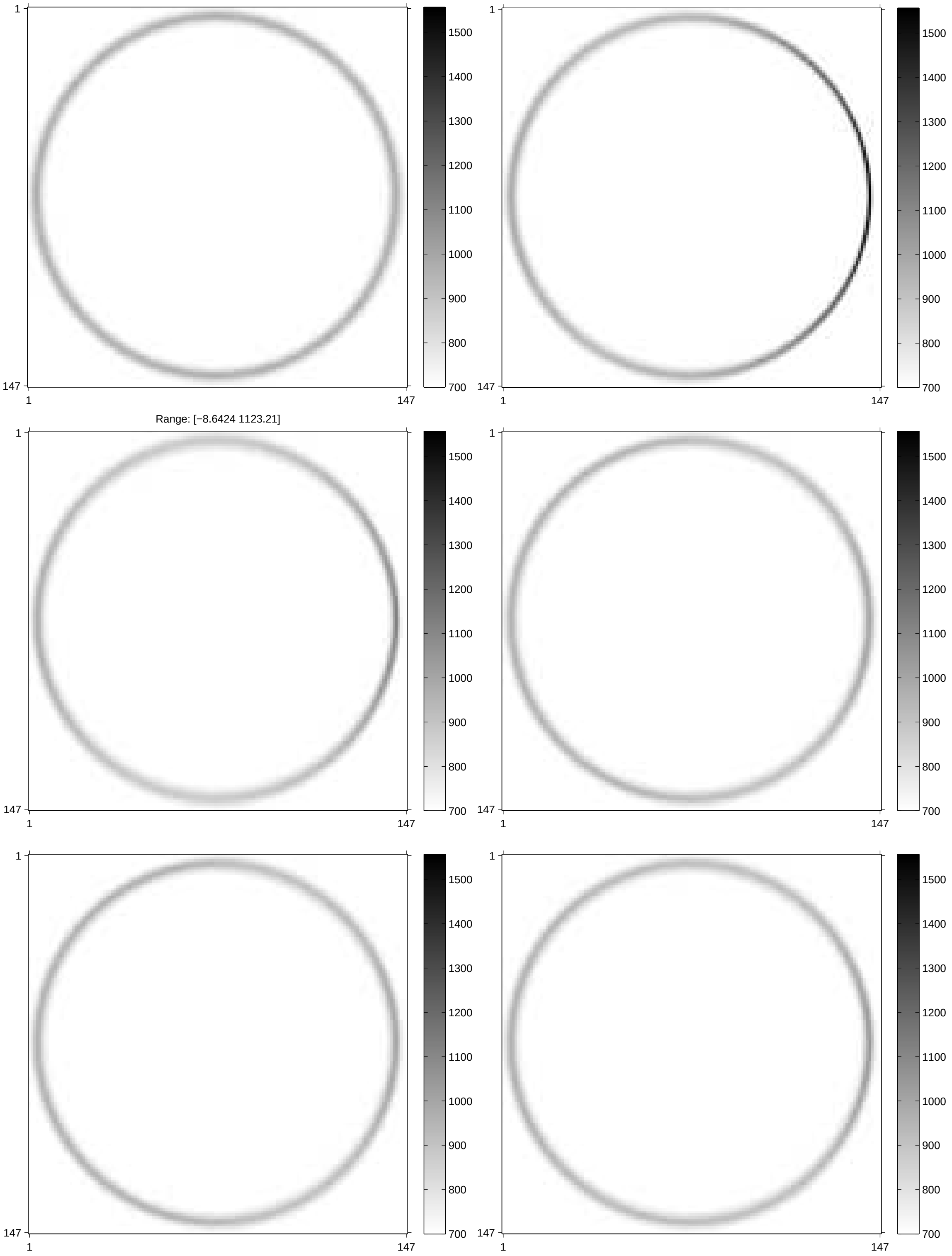


Figure 5.3: Images of right-most ring, Upper-Left: uniformly blurred by target PSF. Upper-Right: reconstructed using conventional regularization. Mid-Left: reconstructed using certainty-based regularization. Mid-Right: reconstructed using *AIMA* regularization, with $\alpha = 0.1$. Lower-Left: reconstructed using *AIMA* regularization, with $\alpha = 0$. Lower-Right: reconstructed using *FIIN* regularization, with $\alpha = 0$.

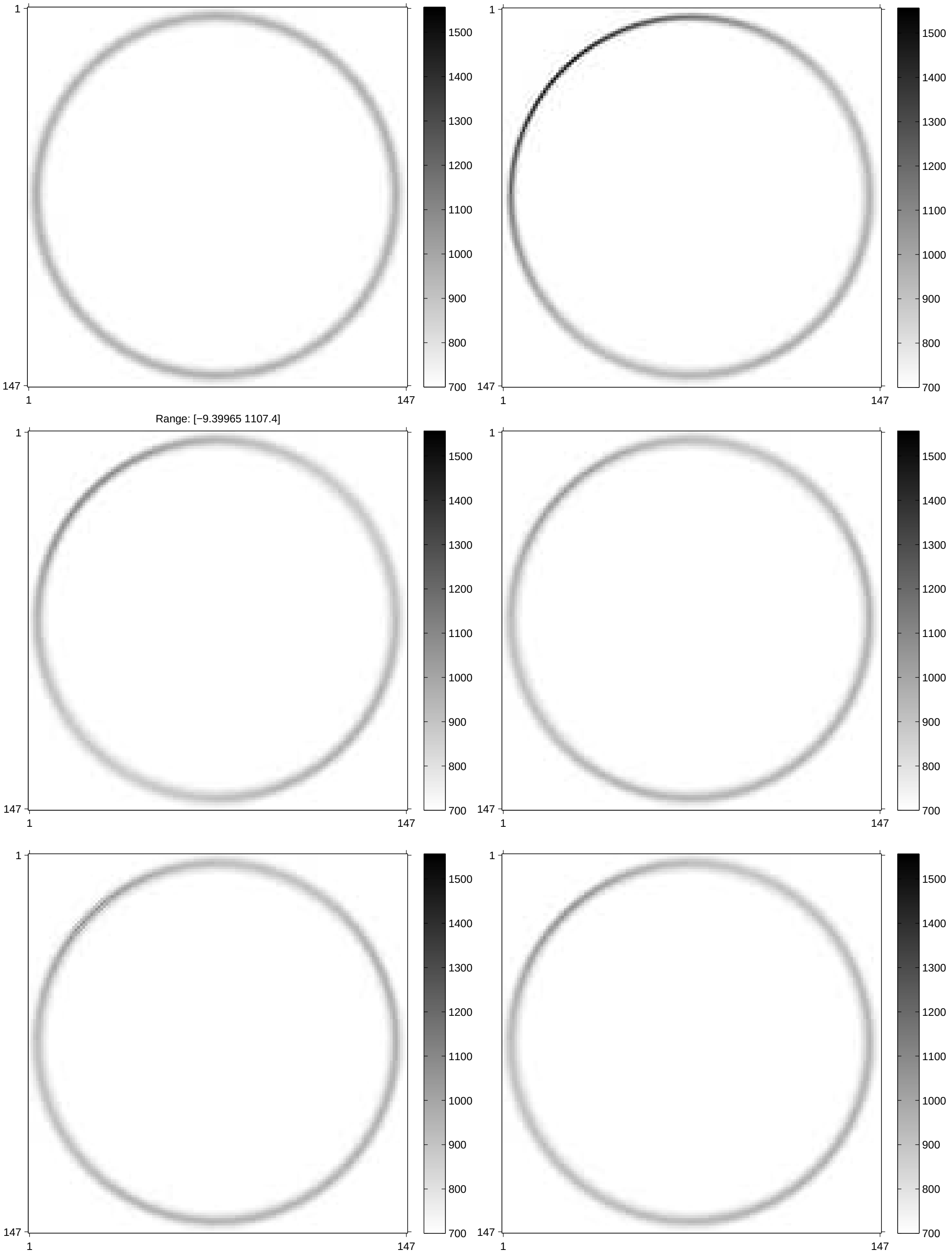


Figure 5.4: Images of left-most ring, Upper-Left: uniformly blurred by target PSF. Upper-Right: reconstructed using conventional regularization. Mid-Left: reconstructed using certainty-based regularization. Mid-Right: reconstructed using *AIMA* regularization, with $\alpha = 0.1$. Lower-Left: reconstructed using *AIMA* regularization, with $\alpha = 0$. Lower-Right: reconstructed using *FIIN* regularization, with $\alpha = 0$.

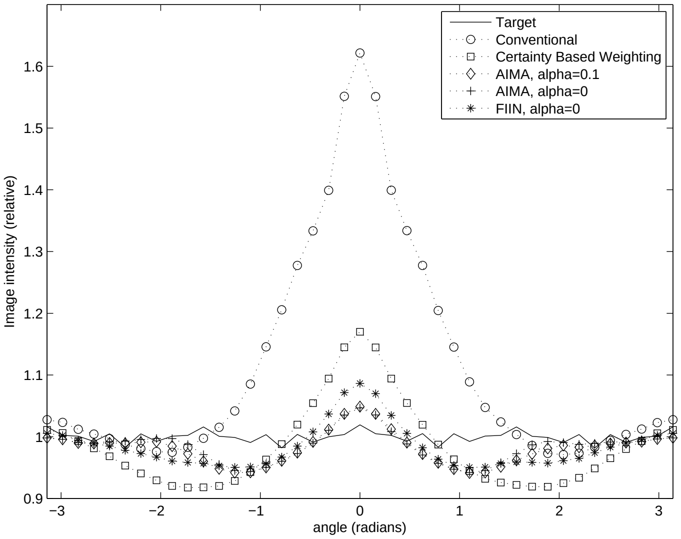


Figure 5.5: Profiles around the right-most ring from each reconstructed image.

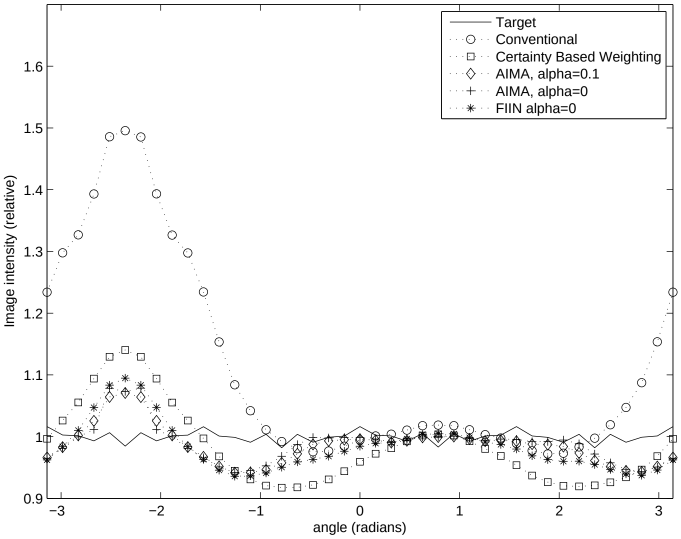


Figure 5.6: Profiles around the left-most ring from each reconstructed image.

Figure 5.7: Reconstruction with conventional regularization without windowing with impulse responses locations marked.

Figure 5.8: Reconstruction with conventional regularization windowed between 800 and 1200 HUs.

Figure 5.9: Reconstruction with certainty based regularization windowed between 800 and 1200 HUs.

Figure 5.10: Reconstruction with *AIMA* regularization with $\alpha = 0.1$ windowed between 800 and 1200 HUs.

Figure 5.11: Reconstruction with *AIMA* regularization with $\alpha = 0$ windowed between 800 and 1200 HUs.

Figure 5.12: Reconstruction with *FIIN* with $\alpha = 0$ windowed between 800 and 1200 HUs.

than the certainty-based approach of [23].

To quantify the performance of these regularizers on real CT data, we computed the PSF of the regularizers at every 10 pixels within the body. Then we calculated the FWHM of the PSFs at 181 evenly spaced angles, and computed the RMS of the error between the actual FWHM and the FWHM of the target. Histograms of the RMS errors are displayed in Fig. 5.17. The mean of the RMS errors for conventional regularization, certainty based regularization, the *AIMA* method with $\alpha = 0.1$, the *AIMA* method with $\alpha = 0$, and the *FIIN* method with $\alpha = 0$ are 2.7, 2.7, 2.3, 2.5, and 2.0 respectively.

Figure 5.13: Impulse Responses at (-100,-100). From left to right, target, conventional regularization, certainty based regularization, *AIMA* regularization with $\alpha = 0.1$, *AIMA* regularization with $\alpha = 0$, *FIIN* regularization with $\alpha = 0$.

5.1.3 Spatial Resolution Properties

Fig. 5.3 and Fig. 5.4 provide a qualitative understanding of the spatial resolution properties of various regularization methods for the ring phantom. The phantom used consists of rings of uniform intensity and uniform width, thus images with uniform spatial resolution

Figure 5.14: Impulse Responses at (150,-120). From left to right, target, conventional regularization, certainty based regularization, *AIMA* regularization with $\alpha = 0.1$, *AIMA* regularization with $\alpha = 0$, *FIIN* regularization with $\alpha = 0$.

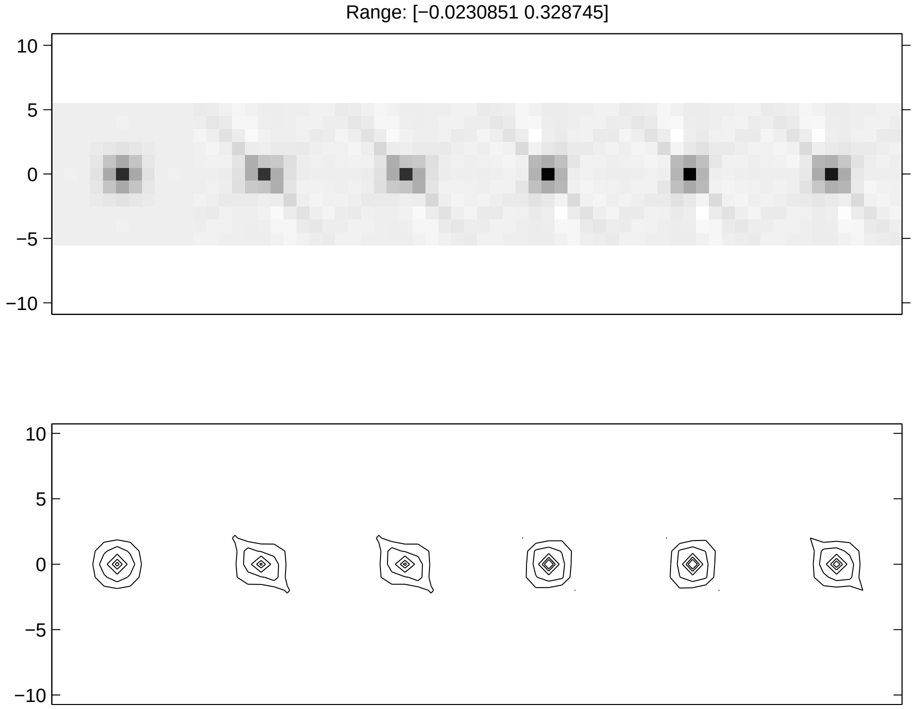


Figure 5.15: Impulse Responses at (170,0). From left to right, target, conventional regularization, certainty based regularization, *AIMA* regularization with $\alpha = 0.1$, *AIMA* regularization with $\alpha = 0$, *FIIN* regularization with $\alpha = 0$.

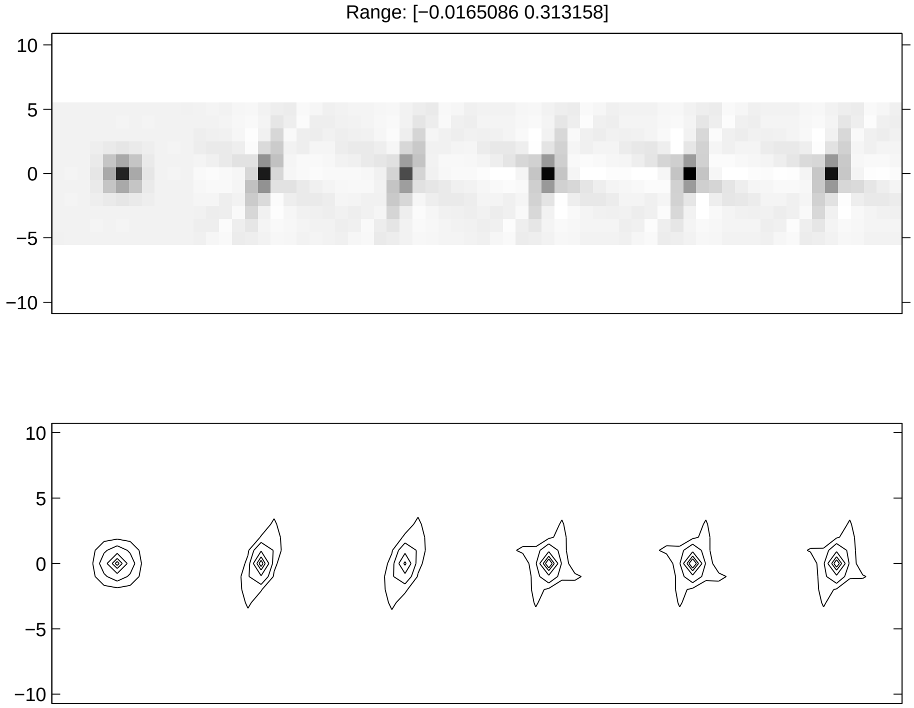


Figure 5.16: Impulse Responses at (-130,100). From left to right, target, conventional regularization, certainty based regularization, *AIMA* regularization with $\alpha = 0.1$, *AIMA* regularization with $\alpha = 0$, *FIIN* regularization with $\alpha = 0$.

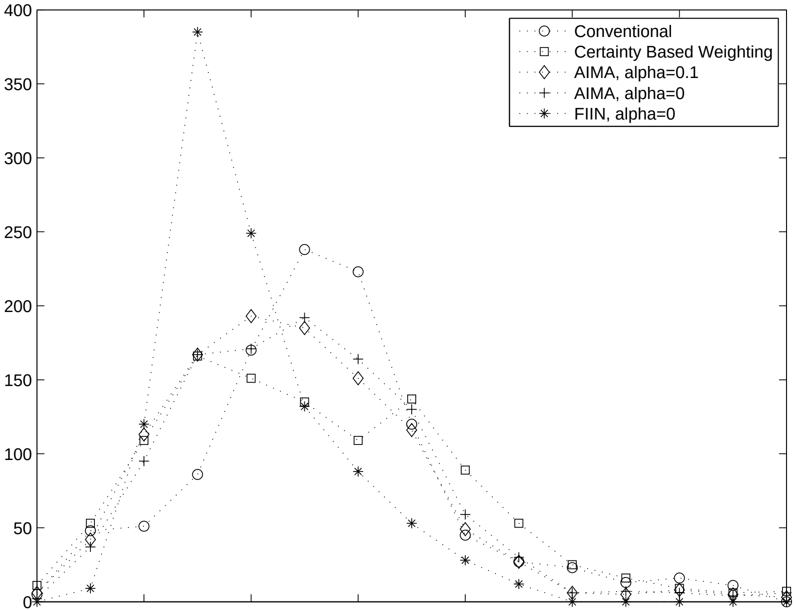


Figure 5.17: Plots of the FWHM error histogram for different impulse responses.

should have rings with uniform width and uniform intensity. Conventional regularization creates rings with sharper spatial resolution near the edge of the field of view. Certainty based regularization improves ring uniformity. The last three reconstructions using *AIMA* with $\alpha = 0.1$, *AIMA* with $\alpha = 0$ and *FIIN* with $\alpha = 0$ provide rings that look almost identical, and that have a more uniform ring width than conventional, and certainty based regularization.

Fig. 5.5 and Fig. 5.6 show the amplitude of the rings traced clockwise. This confirms our initial assessment, that certainty based regularization outperforms conventional regularization, and that the *AIMA* and *FIIN* methods have similar spatial resolution properties and outperform the previous approaches.

Fig. 5.8 - Fig. 5.12 display a quadrant of windowed reconstructions using real CT data to illustrate the images produced using these regularization design methodologies. However, since we do not know the “truth” for this data, these images provide only a qualitative understanding of the effect of regularization design on spatial resolution. The impulse responses in Fig. 5.13-Fig. 5.16 illustrate the effect of regularization design on spatial resolution at various locations. These figures confirm that *AIMA* and *FIIN* methods improved isotropy over conventional and certainty-based regularization. The histogram plot of Fig. 5.17 and the mean of the PSF errors mentioned previously confirm this. *AIMA* outperforms both certainty based regularization and conventional regularization. *FIIN* outperforms all other regularization designs in terms of the FWHM of the local PSFs but is much slower than *AIMA*.

The resulting impulse responses from the *AIMA* and *FIIN* methods are not completely isotropic. This may seem to contradict the dramatic improvement these regularization design methods achieved with the ring phantom. However, recall that we are trying to approximate $\tilde{w}^j(\Phi)$ using 3 basis functions (see §4.1.2) for the *AIMA*, and 4 basis functions (see §4.1.1) for *FIIN*. With real data, $\tilde{w}^j(\Phi)$ is a complicated function that cannot be parameterized using 3 or 4 basis functions. That aspect, along with the non-negativity constraint limits the performance of any regularization design technique with a finite number of parameters. $\tilde{w}^j(\Phi)$ is much simpler for simple phantoms like the ring phantom, so *AIMA* yields better results there. Extensions of this regularization design to higher order penalties have the potential for more basis functions, and better performance.

The analysis of this paper focuses on the resolution nonuniformities cased by statistical weightings, not the resolution variation due to detector response and magnification. A more general regularization design with similar parameterization is discussed in [67]. Using the techniques in this paper to account for these effects is an open problem.

5.1.4 Computation Time

AIMA is quite efficient. Computing certainty based regularization takes the time of about 1 backprojection. In *AIMA*, we must first compute $\tilde{w}^j(\Phi)$ which takes the time of about 1 backprojection, and then solve the analytical solution which is very fast. *FIIN* also requires 1 backprojection to compute $\tilde{w}^j(\Phi)$, however it then has to run a non-negative least squares problem for every pixel. Though this is a fairly small NNLS problem, T is 4×4, it adds much compute time since it must be calculated for each pixel. In general, due to the faster compute times, we recommend *AIMA* with $\alpha = 0.1$. If accuracy is more important than compute time, *FIIN* can be used instead.

5.2 3D Results

5.2.1 Spherical Shells

We simulated a 3D PET system that images a 400mm x 400mm x 164mm volume. The simulated imaging system has 143 rays per view spaced 4mm apart, 80 evenly spaced view angles rotated along Φ, and 5 cross plane angles for Θ. The reconstructed images consist of 100x100x41 4mm pixels. We simulated a noiseless sinogram using our true image which consists of 2 spherical shells inside a cylinder. We choose a target spatial resolution with a 2 pixel FWHM, or equivalently 4 mm.

Fig. 5.18 and Fig. 5.19 show cross sections of 3D PSFs at different locations, showing the xy slice, xz slice, and yz slice from top to bottom. From left to right, are the PSFs at (25,25,0), (15,15,0), (15,15,15), and (25,0,0). PSFs from conventional regularization are shown in Fig. 5.18. PSFs from *AIMA* regularization are shown in Fig. 5.19. Looking at these images, one may observe a great deal of anisotropy in the PSFs created from conventional regularization. The xy slices are stretched out in different directions depending on where the pixel lies with respect to the center of the image. The xz and yz slices all seem flatter and smaller when compared to xy slices. Looking at the PSFs generated using *AIMA* regularization, each slice looks more isotropic, and each slice looks to be about the same size. The flatness of xz and yz slices has been removed, and the xy slices are no longer stretched. The PSF at (25,25,0) displays some anisotropy. The xy slice appears to be boxy with tails, instead of being round, and the xz and yz slices are not completely isotropic. This shows the limitations of this method in approximating $w^j(\Phi)$.

In the next few figures, we take xy, xz, and yz slices of 2 PSFs, and then measure the FWHM cutting through the center of the PSF at different angles to get a sense of isotropy. The uniformity of these plots are a rough measure of the isotropy of these impulse functions. Figure 5.20 and 5.21 shows plots of the FWHM for PSFs at (25,25,0) and (15,15,15).

Tables 5.1 displays maximum and minimum FWHMs for PSFs at (25,25,0) and (15,15,15) allowing for a more quantitative assessment of isotropy. Even for (25,25,0), which exhibits less improvement than (15,15,15), there is less variation in FWHM for PSFs generated using *AIMA* regularization than PSFs generated using conventional regularization.

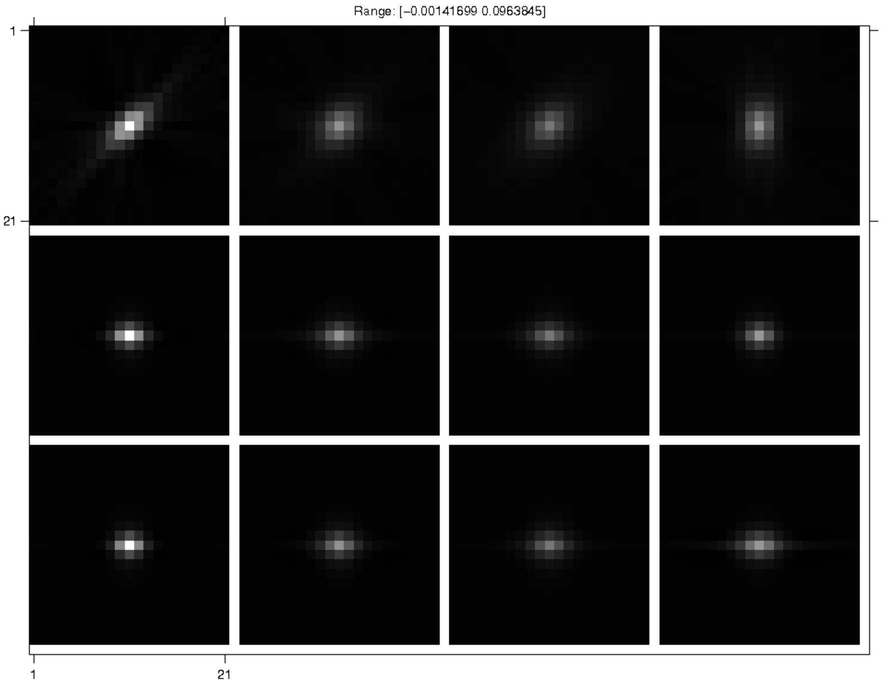


Figure 5.18: xy, xz, and yz (top to bottom) slices of PSFs created using conventional regularization at (25,25,0), (15,15,0), (15,15,15), and (25,0,0) (left to right).

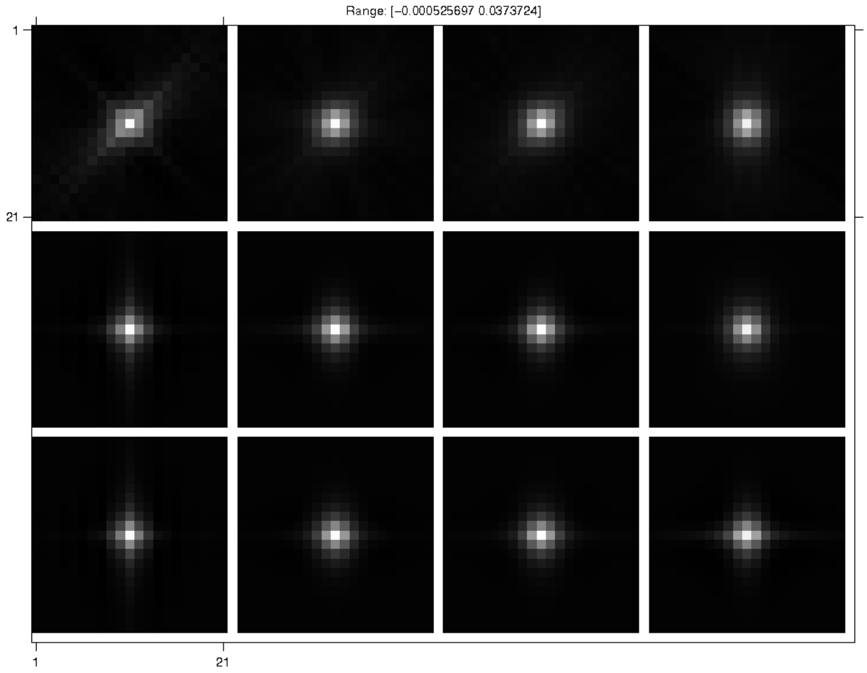


Figure 5.19: xy, xz, and yz (top to bottom) slices of PSFs created using *AIMA* regularization at (25,25,0), (15,15,0), (15,15,15), and (25,0,0) (left to right).

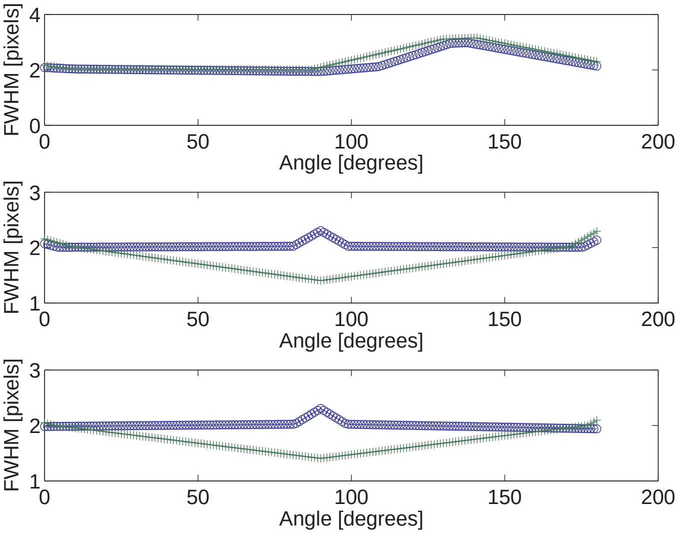


Figure 5.20: FWHM plots for PSF at (25,25,0). *AIMA* regularization marked with circles, conventional marked with pluses. xy, xz, yz slices from top to bottom.

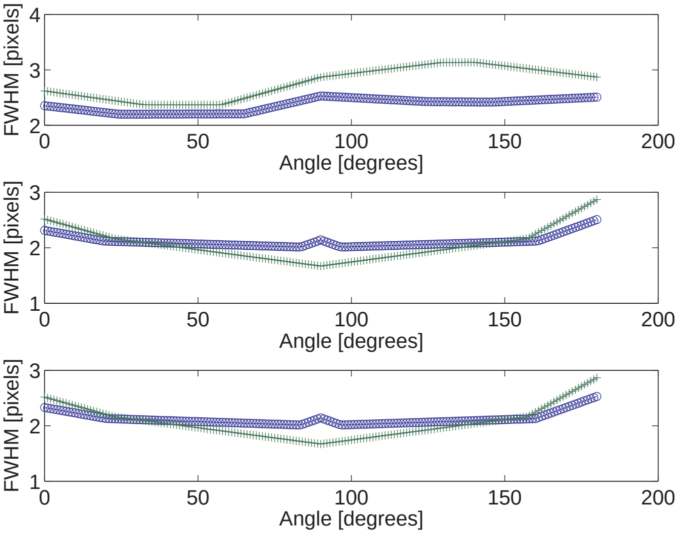


Figure 5.21: FWHM plots for PSF at (15,15,15). *AIMA* regularization marked with circles, conventional marked with pluses. xy, xz, yz slices from top to bottom.

Table 5.1: Maximum and Minimum FWHM values for 2 PSFs.

(25,25,0)	*AIMA*		Conventional	
	min	max	min	max
xy	1.9387	2.9770	2.0035	3.1553
xz	2.0061	2.3060	1.4048	2.2946
yz	1.9387	2.3060	1.4048	2.0932
(15,15,15)	*AIMA*		Conventional	
	min	max	min	max
xy	2.1954	2.5281	2.3689	3.1347
xz	2.0137	2.5069	1.6706	2.8698
yz	2.0141	2.5281	1.6706	2.8676

Next we look at reconstructions using various regularizers xy slices of this object are shown in figure 5.22, xz slices of this object are shown in figure 5.23, yz slices of this object are shown in figure 5.24. We reconstructed volumes using an incremental optimization transfer version of the EM algorithm [1] with two types of regularization. We have 2 sets of volumes resulting from conventional regularization, and *AIMA* regularization. We show xy slices 11, 13, 17, 21, and 27 of the volume, followed by xz slices 36, 41, 46, 51, 56, 61, followed by yz slices 21, 31, 41, 61, 71, 81. Volumes reconstructed using conventional regularization are shown in Fig. 5.25, Fig. 5.26, and Fig. 5.27. Volumes reconstructed using *AIMA* regularization are shown in Fig. 5.28, Fig. 5.29, and Fig. 5.30. Volumes reconstructed using *AIMA* regularization design show greater uniformity than those reconstructed using conventional regularization.

Finally, we simulated a noisy sinogram. Volumes reconstructed using conventional regularization are shown in figures 5.31, 5.32, and 5.33. Volumes reconstructed using *AIMA* regularization design are shown in figures 5.34, 5.35, and 5.36. Although it is less apparent due to the noise in the images, Volumes reconstructed using *AIMA* regularization still show greater uniformity than those reconstructed using conventional regularization.

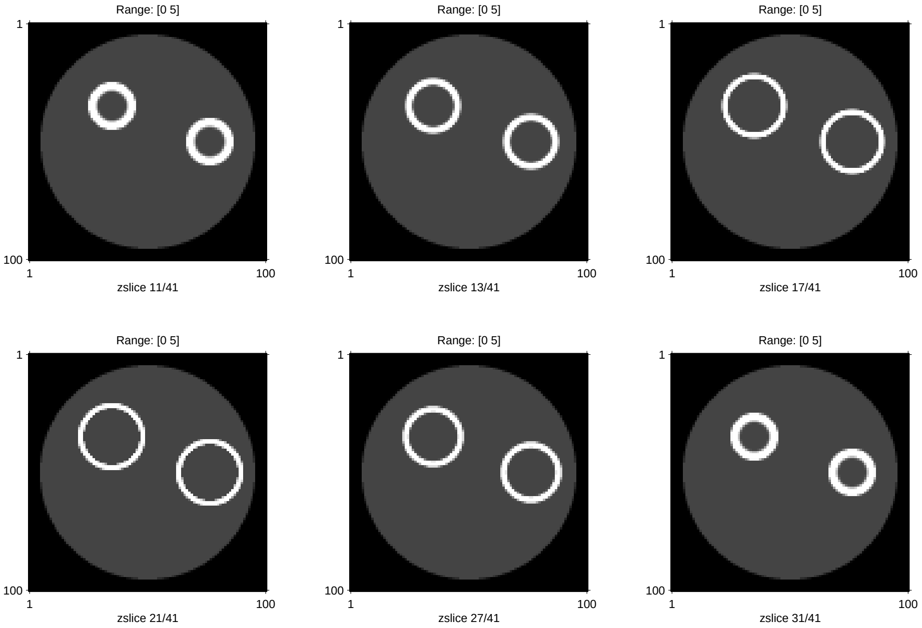


Figure 5.22: xy slices of the original object.

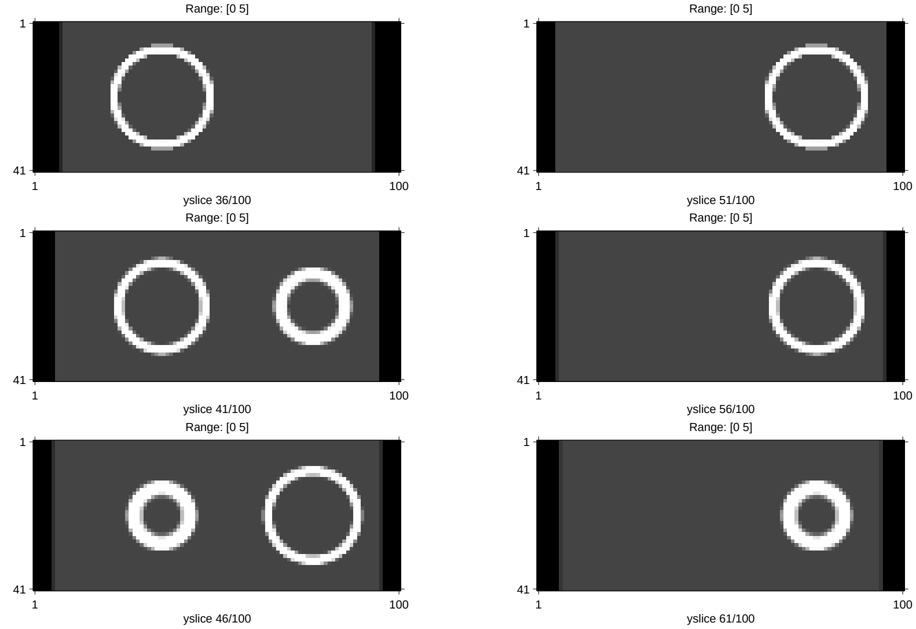


Figure 5.23: xz slices of the original object.

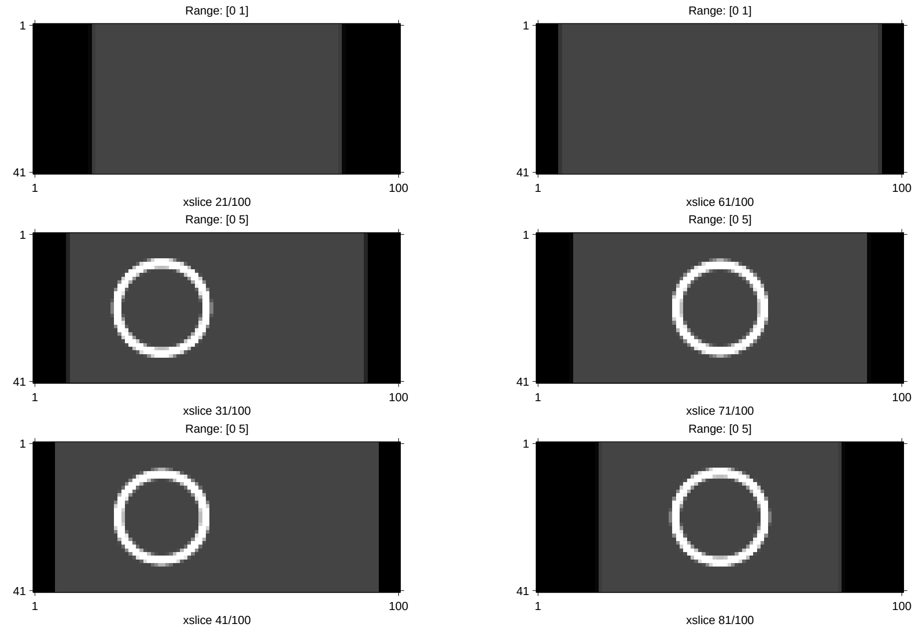


Figure 5.24: yz slices of the original object.

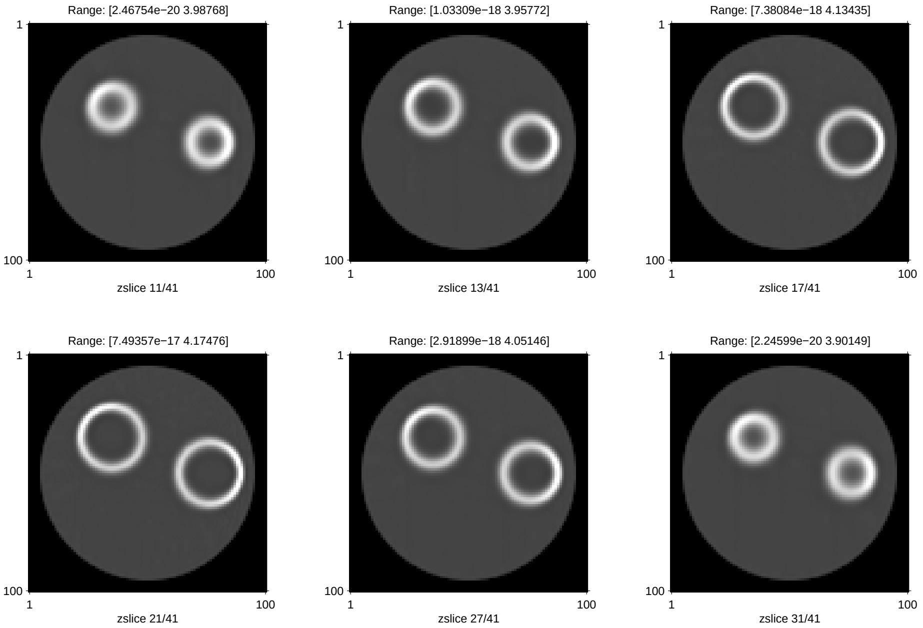


Figure 5.25: xy slices of PL reconstructed images with conventional regularization design from noiseless data.

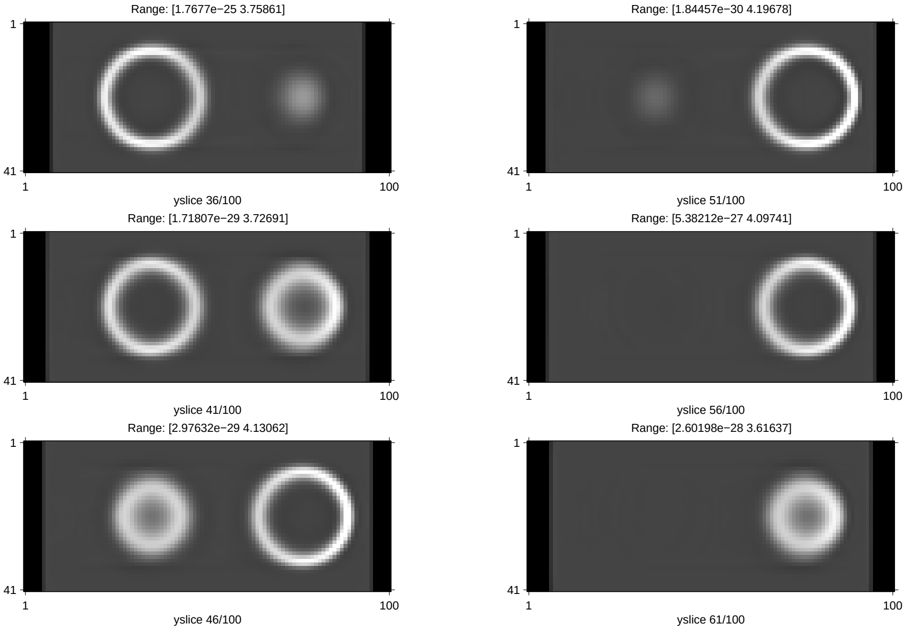


Figure 5.26: xz slices of PL reconstructed images with conventional regularization design from noiseless data.

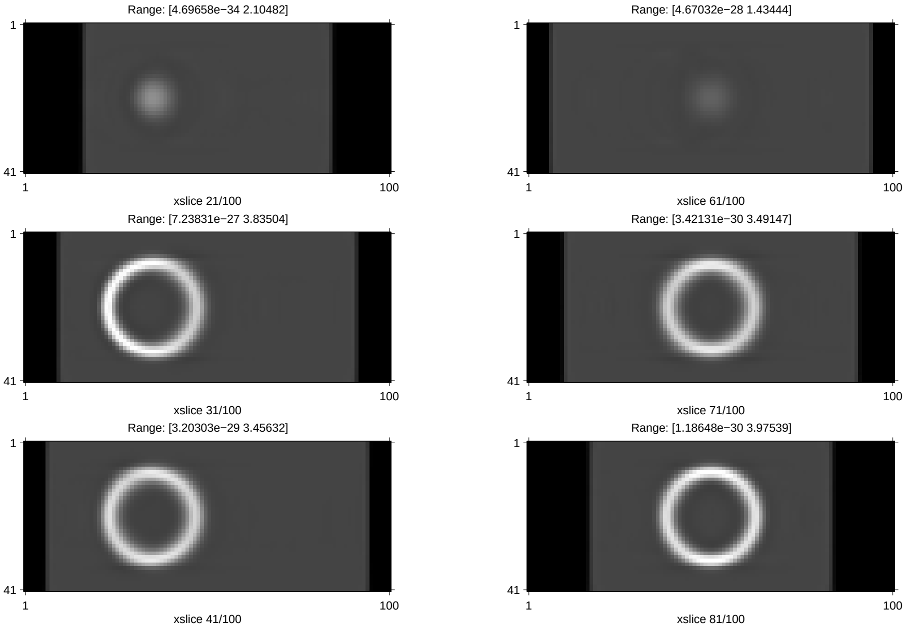


Figure 5.27: yz slices of PL reconstructed images with conventional regularization design from noiseless data.

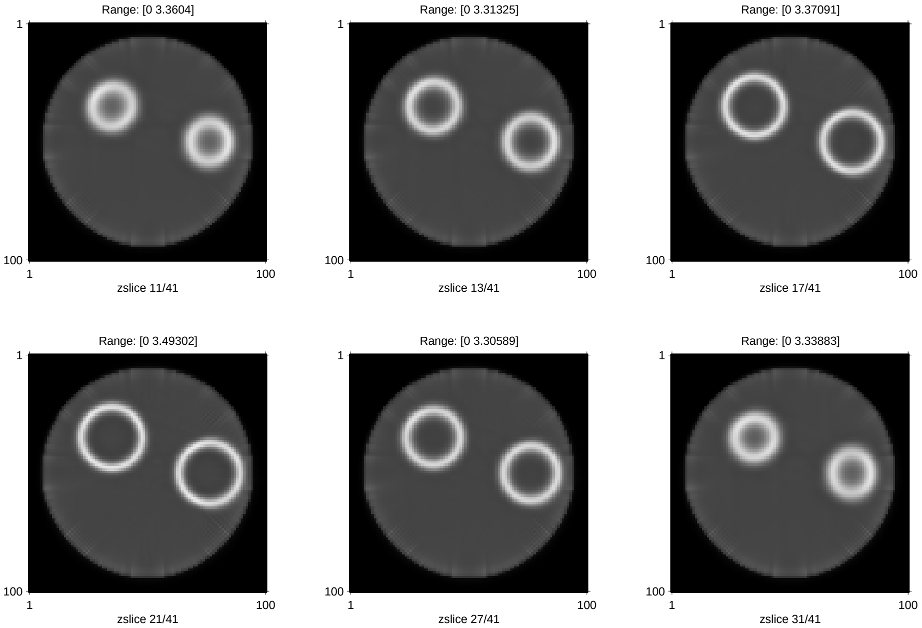


Figure 5.28: xy slices of PL reconstructed images with *AIMA* regularization from noiseless data.

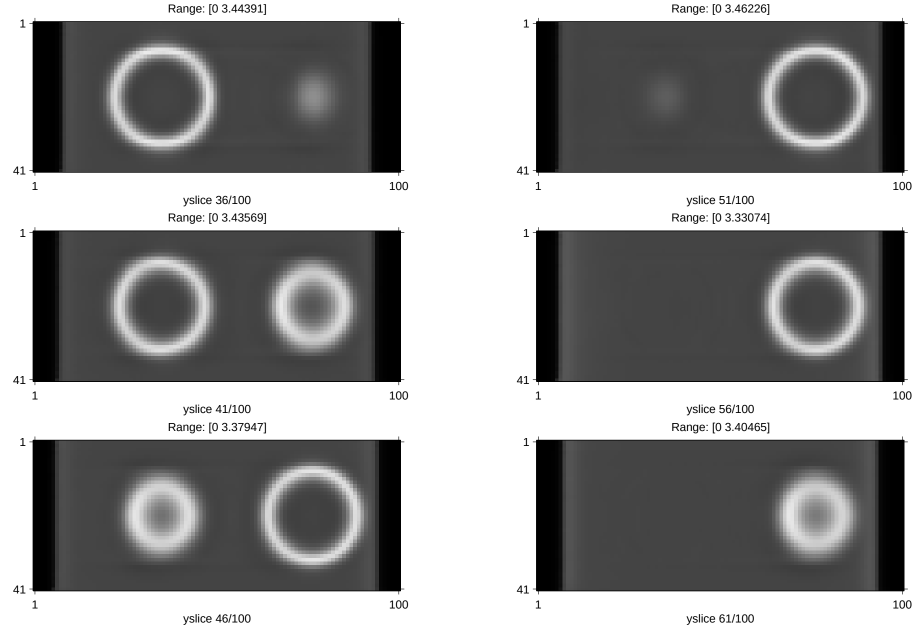


Figure 5.29: xz slices of PL reconstructed images with *AIMA* regularization from noiseless data.

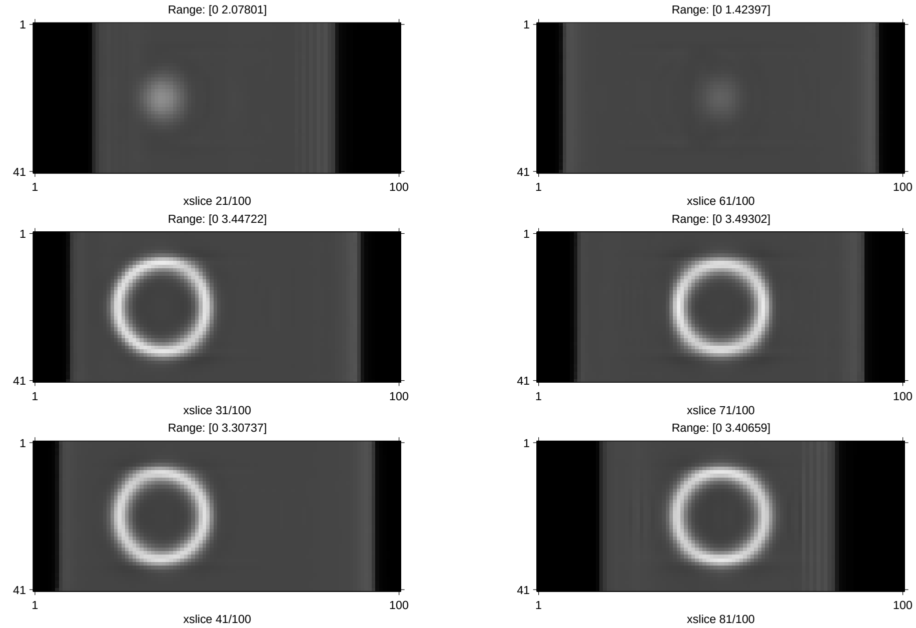


Figure 5.30: yz slices of PL reconstructed images with *AIMA* regularization from noiseless data.

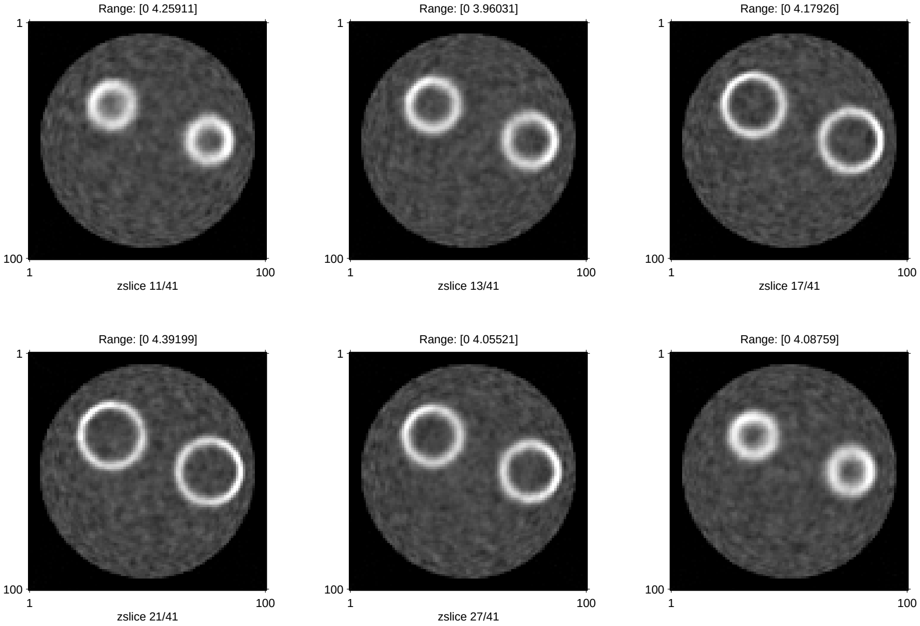


Figure 5.31: xy slices of PL reconstructed images with conventional regularization design from noisy data.

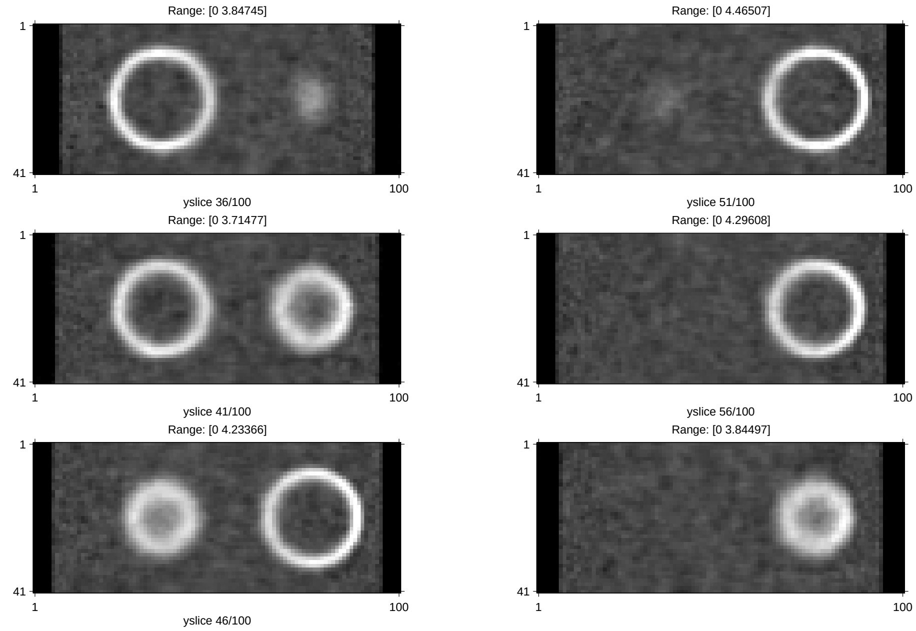


Figure 5.32: xz slices of PL reconstructed images with conventional regularization design from noise data.

5.2.2 Zubal Phantom

We next use the same system model to compute PSFs using the Zubal phantom [44,78], to get an idea of how 3D regularization performs under more realistic conditions. Unlike the ring phantom, which makes anisotropy and non-uniformity more obvious, the effects on the Zubal phantom are more subtle. We focus our analysis on PSFs to gain insight into the resolution properties of 3D regularization. Fig. 5.37-5.40 Show XY, XZ, and YZ slices of PSFs at various locations. PSFs were computed using certainty based regularization [23], *AIMA* with $\alpha = 0$, and *FIIN* with $\alpha = 0$. PSFs constructed from 2D regularization have also been shown as a point for comparison. In these figures, ζ was tuned so that the FWHM was around 2 pixels, or 8 mm

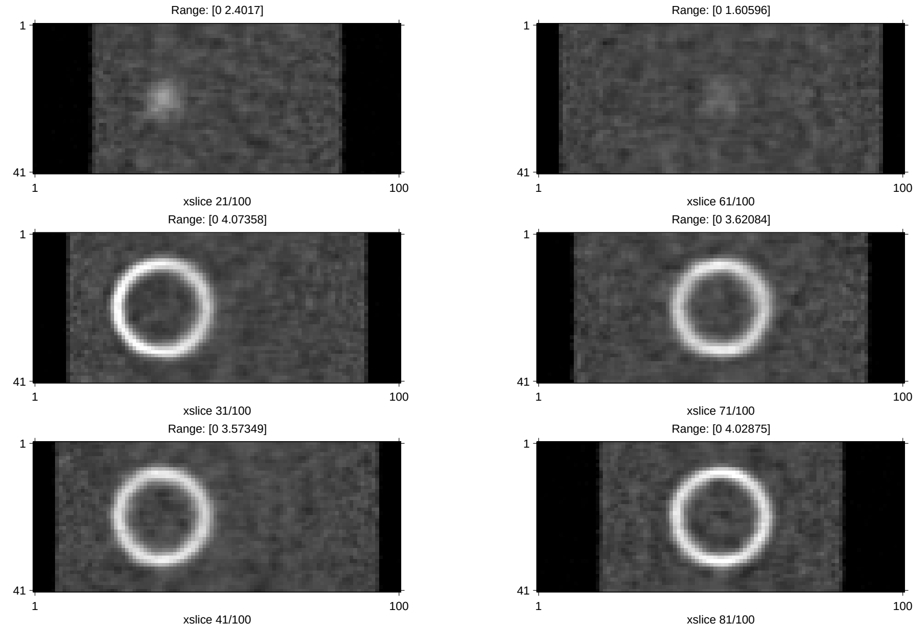


Figure 5.33: yz slices of PL reconstructed images with conventional regularization design from noisy data.

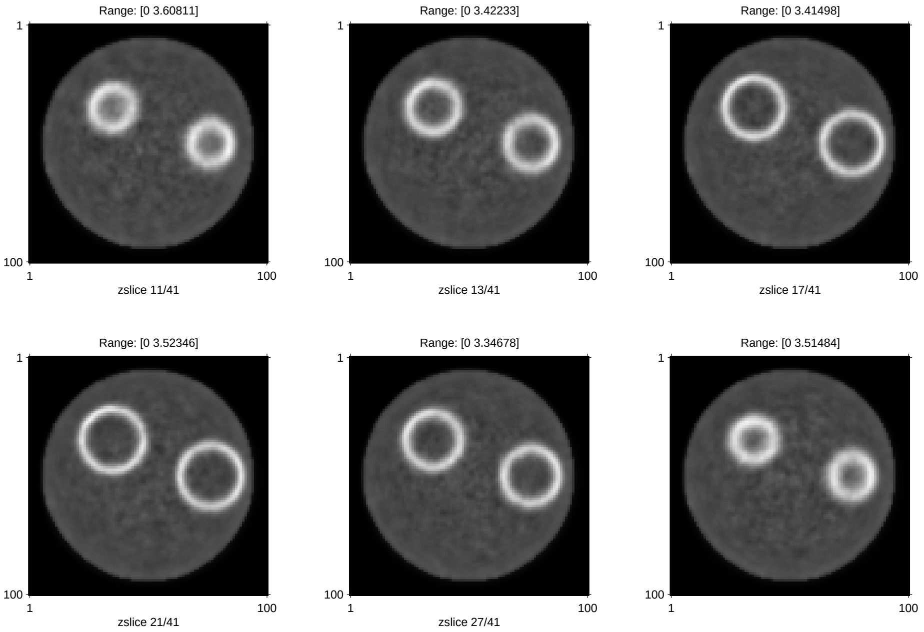


Figure 5.34: xy slices of PL reconstructed images with *AIMA* regularization from noisy data.

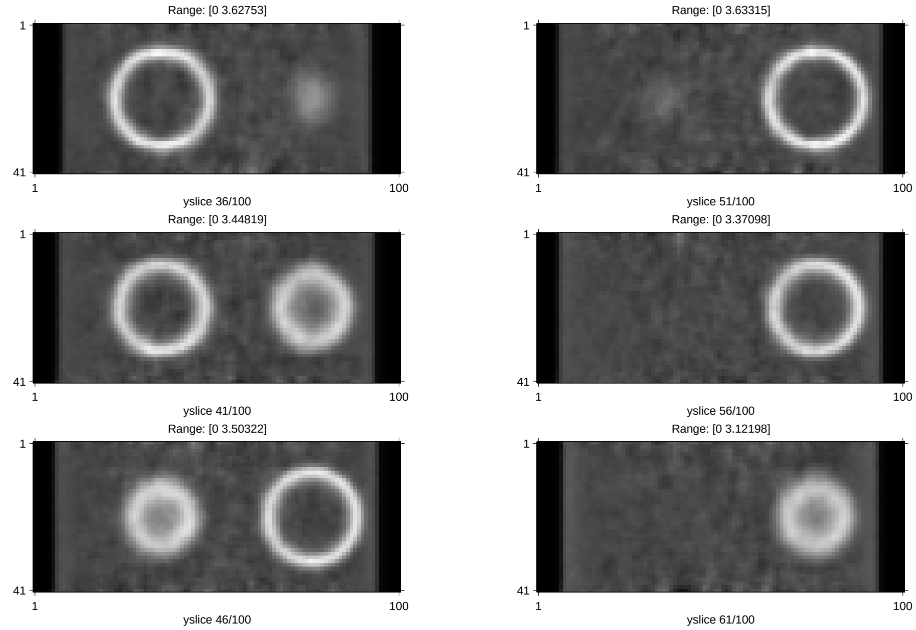
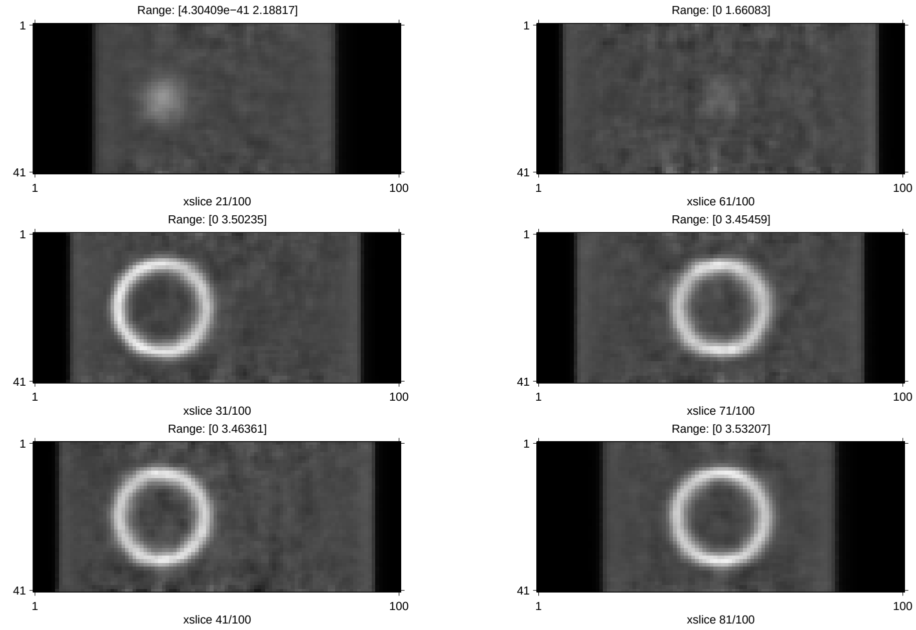


Figure 5.35: xz slices of PL reconstructed images with *AIMA* regularization from noisy data.


Range: [4.30409e−41 2.18817]

xslice 21/100

Range: [0 1.66083]

xslice 61/100

Range: [0 3.50235]

xslice 31/100

Range: [0 3.45459]

xslice 71/100

Range: [0 3.46361]

xslice 41/100

Range: [0 3.53207]

xslice 81/100

Figure 5.36: yz slices of PL reconstructed images with *AIMA* regularization from noisy data.

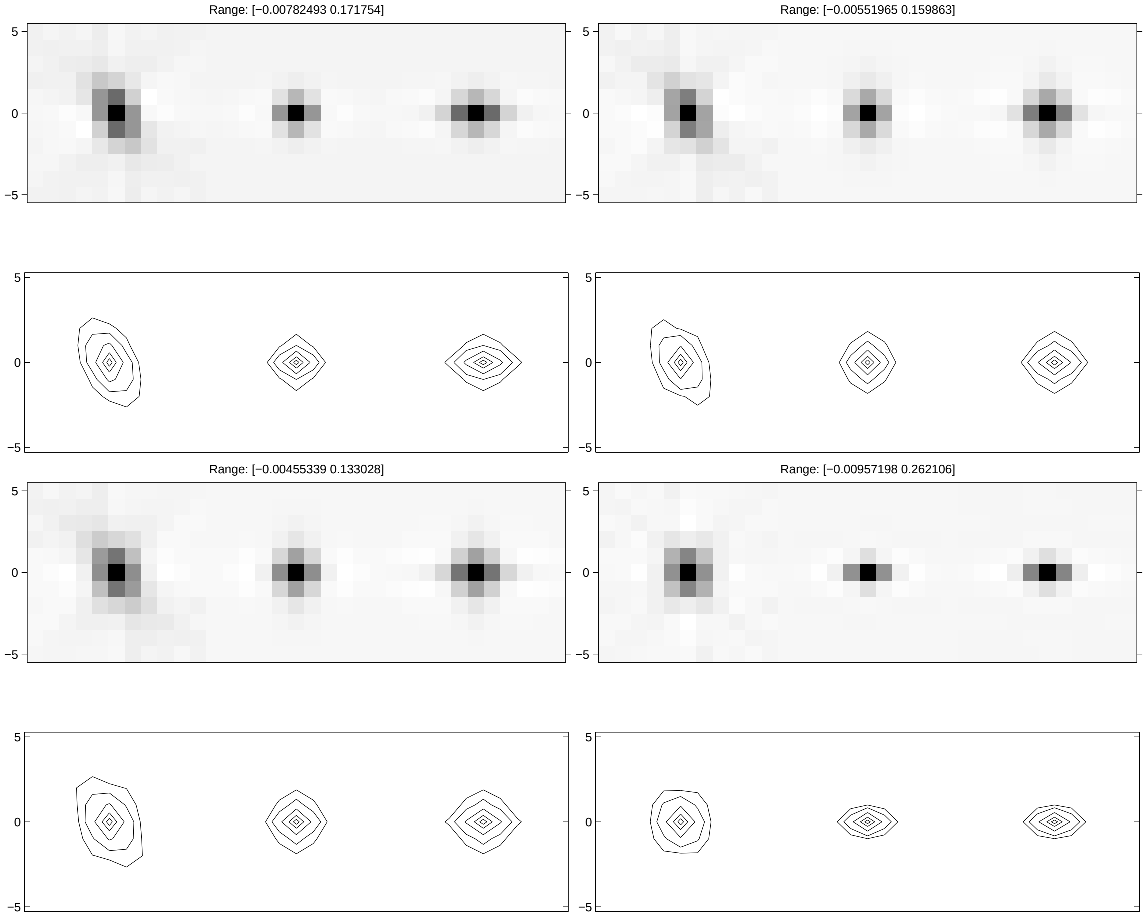


Figure 5.37: PSF at (30,0,0), using, top-left, certainty based regularization, top-right, *AIMA* $\alpha = 0$, bottom left, *FIIN*, bottom right, 2D regularization. Each plot shows from left to right, xy, xz, and yz slices through the PSF.

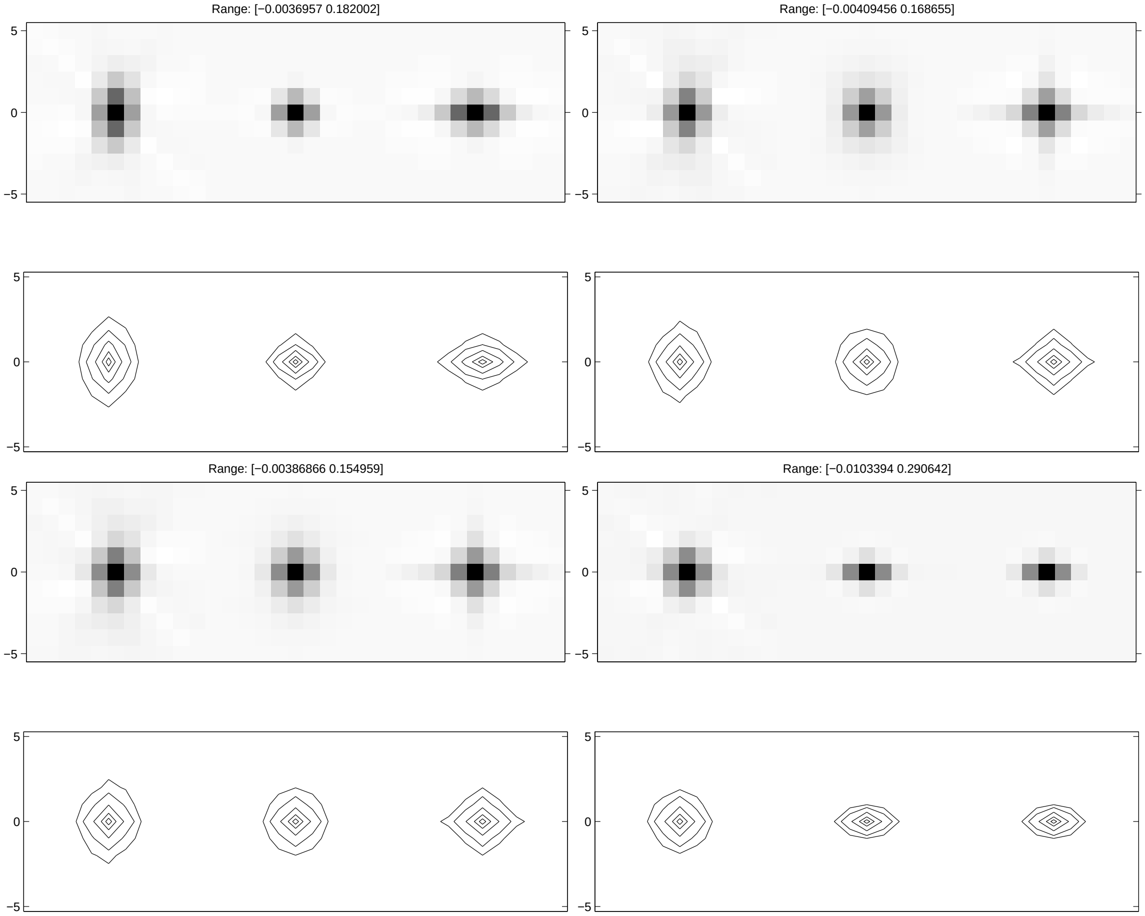


Figure 5.38: PSF at (-15,-15,0), using, top-left, certainty based regularization, top-right, *AIMA* $\alpha = 0$, bottom left, *FIIN*, bottom right, 2D regularization. Each plot shows from left to right, xy, xz, and yz slices through the PSF.

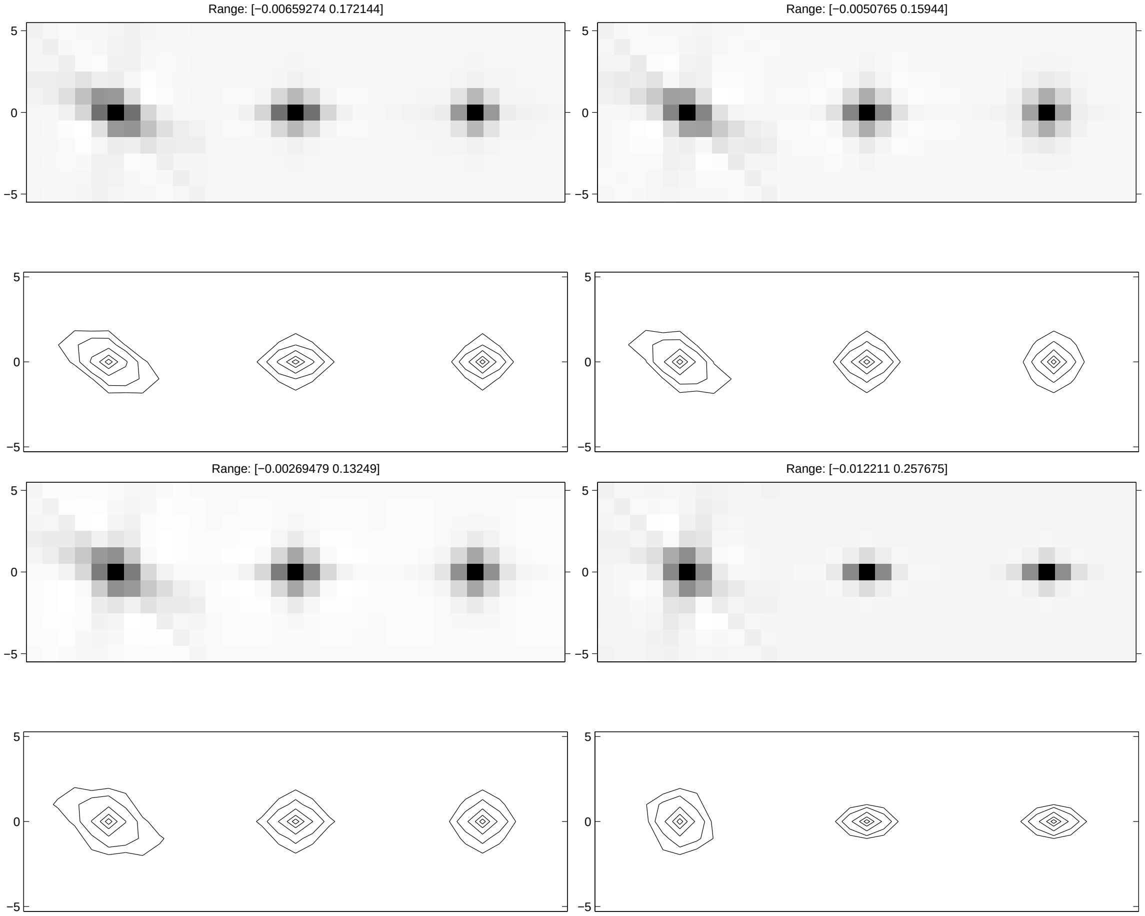
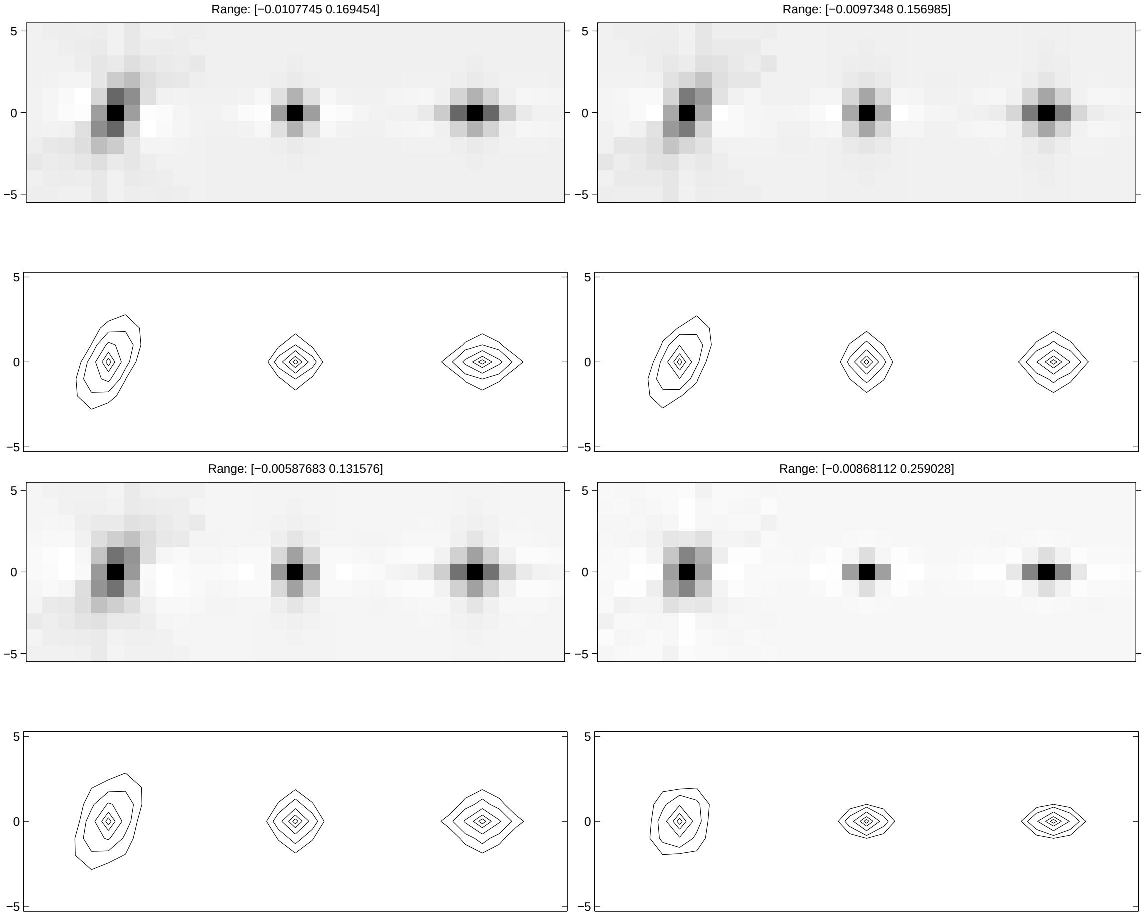


Figure 5.39: PSF at (15,15,-10), using, top-left, certainty based regularization, top-right, *AIMA* $\alpha = 0$, bottom left, *FIIN*, bottom right, 2D regularization. Each plot shows from left to right, xy, xz, and yz slices through the PSF.


Range: [−0.0107745 0.169454]

Range: [−0.0097348 0.156985]

Range: [−0.00587683 0.131576]

Range: [−0.00868112 0.259028]

Figure 5.40: PSF at (-30,0,10), using, top-left, certainty based regularization, top-right, *AIMA* $\alpha = 0$, bottom left, *FIIN*, bottom right, 2D regularization. Each plot shows from left to right, xy, xz, and yz slices through the PSF.

5.2.3 3D results discussion

This section discusses the results of 3D regularization. Our results show that using the *AIMA* method of regularization design can correct for some anisotropy, however the PSFs associated with 3D regularization appear less isotropic in the XY plane than 2D regularization.

5.2.4 Spatial Resolution

3D regularization performs well for the phantom with spherical shells. The images of the PSFs as well as the reconstructed volume provide a qualitative understanding of how 3D regularization using the *AIMA* method can improve spatial resolution. Plots of the FWHM in Fig. 5.20 and Fig. 5.21 for PSFs at two locations provide quantitative evidence that *AIMA* can improve resolution isotropy for the spherical shell phantom. Reconstructed images show that 3D regularization is also successful in the presence of noise.

3D regularization improves resolution isotropy for the Zubal phantom. However, like in 2D, when we move from rings or spherical shells to more realistic data, the improvement of 3D regularization is less dramatic. The *AIMA* method and the *FIIN* method appear to perform equally well, and both outperform certainty based regularization. However, 2D regularization provides for the greatest improvement in XY-plane isotropy. This is discussed at length in §VI.

5.2.5 Computational Performance

AIMA is much slower in 3D than in 2D because there does not appear to be an analytical solution to the minimization problem in (4.18). This means that an iterative NNLS algorithm must be run for every voxel. Coupled with the fact that 3D volumes have many more voxels than 2D pixels, 3D regularization may be impractical. Furthermore, 3D regularization requires storing L coefficients per voxel. With smaller image dimensions in

PET, this may be feasible, but in CT, with up to 13 neighbors, the memory requirements become unmanageable. Since the regularization coefficients change slowly over space, one could use coarser sampling in storing and computing $\boldsymbol{r}^j$. An alternative approach will be discussed in §VI.

CHAPTER VI

Z-directional Post Filtering

As mentioned in the previous chapter, 2D regularization seems to outperform 3D regularization in terms of XY isotropy. Recall the $\cos\Theta$ term in (3.34). This is a mathematical representation of a fundamental anisotropy in 2.5D PET and CT systems, that the inherent blur of these systems occurs mostly in the xy plane. No regularizers can completely compensate for this because it requires a $\cos\Theta$ term in the denominator in the local frequency response, resulting in an infinite response at $\Theta = \pi/2$. We have noticed that in designing 3D regularizers, sometimes we sacrifice xy plane isotropy to try to achieve additional isotropy in the z direction. Fig. 6.1 and Fig. 6.2 show impulse responses resulting from 2D and 3D *AIMA* regularization respectively. These results were generated using all 8 slices of the GE scanner and data mentioned in §5.1.   From these images we can see that 3D

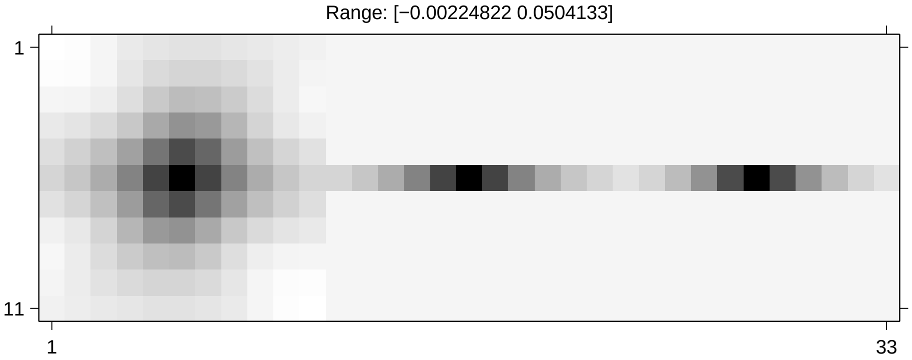


Figure 6.1: Impulse response resulting from *AIMA* 2D regularization, from left to right, xy slice, xz slice, yz slice.

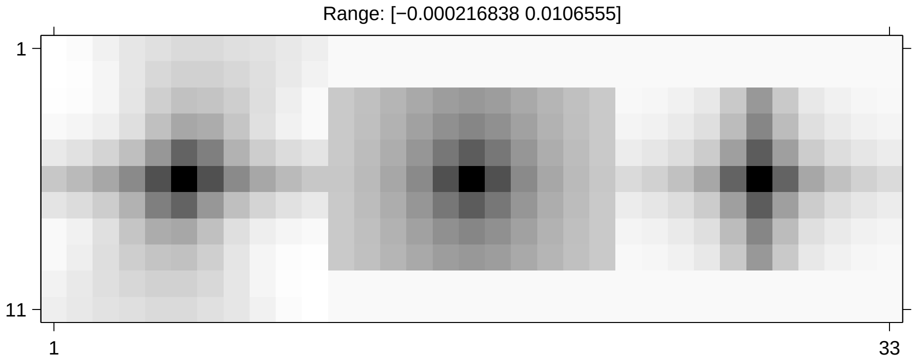


Figure 6.2: Impulse response resulting from *AIMA* 3D regularization, from left to right, xy slice, xz slice, yz slice.

regularization is more isotropic in all 3 directions, however 2D regularization performs better in the XY plane. In addition, 3D regularization requires large amounts of memory, up to 13 regularization coefficients for each pixel. Since scanner size in terms of number of pixels tends to outpace computer memory, these memory issues are very relevant. This chapter discusses the use of 2D regularization, which is more isotropic in the XY plane, and more memory efficient, followed by applying z-dimension post-reconstruction denoising that leads to an memory and computationally efficient solution for 3D isotropy.

We analyze three methods in this chapter, case 1: 3D regularization, case 2: 2D regularization with z-dimension post-reconstruction denoising, and case 3: no regularization with 3D post-reconstruction denoising. We analyze the spatial resolution properties of these three methods, and then the noise properties. Finally we look at reconstructed volumes. In CT typically we wish to use edge-preserving regularization to preserve detail in the reconstructed volume. Some have argued that the analysis for certainty based regularization can be extended to the non-quadratic case [2]. Since this work is related to certainty based weighting, regularization coefficients designed using *AIMA* and *FIIN* may also help achieve uniform and isotropic spatial resolution even with non-quadratic regularization. Unfortunately non-quadratic edge-preserving regularization is difficult to analyze, so we will provide analysis using quadratic regularizers to gain insight in to the spatial resolution and noise properties of these methods, but show reconstructions using edge-preserving regularization.

6.1 Spatial Resolution

In this section we want to analyze and isolate the effect of the three regularization schemes independent of the weightings used, so we assume that $\boldsymbol{W} = \boldsymbol{I}$ and focus on parallel-beam geometry instead of fan-beam geometry. We focus our analysis on the quadratic case in this section. The edge-preserving penalties we consider in this chapter use the Huber penalty and operate in either the quadratic region of the potential function or the edge-preserving region. This analysis is directly applicable when in the quadratic region of the Huber penalty. However to analyze the edge-preserving region, a more sophisticated method such as the local perturbation response [2, 3] must be considered. We attempt to make each of these regularization methods match the isotropic target 3D PSF (3.35) discussed in §3.3 that has the frequency response

$$\mathcal{L}^0(\varrho) = \frac{1}{1 + \zeta(2\pi)^2|\varrho|^3}. \tag{6.1}$$

6.1.1 3D Regularization

We consider a conventional 3D regularizer with 3 adjacent neighbors. Since 2.5D tomographic systems are inherently anisotropic regardless of the weightings used, we separate the conventional regularizer R into 2 components, $R_{xy}(\varrho, \Phi, \Theta) = (2\pi\varrho\cos\Theta)^2$ and $R_z(\varrho, \Phi, \Theta) = (2\pi\varrho\sin\Theta)^2$. These correspond to the local frequency response of regularization with the structure in §3.3.1 with $(n_l, m_l, z_l) \in \{(1, 0, 0), (0, 1, 0)\}$ for R_{xy} and $(n_l, m_l, z_l) \in \{(0, 0, 1)\}$ for R_x. This allows us to apply different amounts of smoothing in the z direction to compensate for the inherent isotropy of 2.5D tomographic systems. The frequency response using this regularizer is

$$\begin{aligned}\mathcal{L}_{\text{3D}} &= \frac{\frac{1}{\varrho\cos\Theta}}{\frac{1}{\varrho\cos\Theta}+\zeta_{xy}R_{xy}(\rho,\Phi,\Theta)+\zeta_z R_z(\rho,\Phi,\Theta)}\\ &= \frac{1}{1+\varrho\cos\Theta\zeta_{xy}R_{xy}(\rho,\Phi,\Theta)+\zeta_z R_z(\rho,\Phi,\Theta)}\\ &= \frac{1}{1+\varrho\cos\Theta(\zeta_{xy}(2\pi\varrho\cos\Theta)^2+\zeta_z(2\pi\varrho\sin\Theta)^2)}\\ &= \frac{1}{1+(2\pi)^2\varrho^3\cos\Theta(\zeta_{xy}\cos^2\Theta+\zeta_z\sin^2\Theta)}. \end{aligned} \tag{6.2}$$

We would like to make this isotropic and match the target (6.2). Toward that goal, we can pick ζ_{xy} and ζ_z to minimize $||\cos\Theta(\zeta_{xy}\cos^2\Theta+\zeta_z\sin^2\Theta)-\zeta||^2$. To solve for ζ_{xy} and ζ_z,

$$\begin{aligned}\cos\Theta(\zeta_{xy}\cos^2\Theta+\zeta_z\sin^2\Theta) &\approx \zeta\\ \zeta_{xy}\cos^3\Theta+\zeta_z\sin^2\Theta\cos\Theta &\approx \zeta\\ \cos^3\Theta+\frac{\zeta_z}{\zeta_{xy}}\sin^2\Theta\cos\Theta &\approx \frac{\zeta}{\zeta_{xy}}.\end{aligned}$$

We solve this numerically and the optimal solution is $\zeta_z \approx 3\zeta_{xy}$.

6.1.2 Post-Reconstruction Denoising

No choice of ζ_{xy} and ζ_z will lead to a perfect match with (6.1), so an alternative approach is to first reconstruct an underlying image $\boldsymbol{x}_{\text{rough}}$ and then denoise it. Because in CT typically non-quadratic edge-preserving regulariation is used, we will consider the spatial resolution of an iterative post-reconstruction de-noising of the form

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} ||\boldsymbol{x}-\boldsymbol{x}_{\text{rough}}||^2+\zeta R_{\text{post}}(\boldsymbol{x}), \tag{6.3}$$

where $\boldsymbol{x}_{\text{rough}}$ is the image before the post-reconstruction denoising, which would be the reconstructed image using WLS for case 3, or the reconstructed image using PWLS with 2D regularization for case 2. R_{post} is the edge preserving post-reconstruction regularizer. For quadratic regularization, the frequency response associated with (6.3) is

$$\mathcal{L}_{\text{post}} = \frac{1}{1+\zeta R_{\text{post}}(\varrho,\Phi,\Theta)}. \tag{6.4}$$

Though the analysis of this section focuses on quadratic regularization, we focus on the form (6.3) rather than generic post-filter so that this work is relevant to edge-preserving regularization.

In image denoising, first-order differencing leads to PSFs with heavy tails. First order differencing does not cause such heavy tails in tomographic reconstruction because of the $\frac{1}{\rho}$ term associated with tomographic systems. We use second-order differencing in our post-reconstruction denoising to help us match the PSFs used in tomographic reconstruction. Stayman and Nuyts have argued [39, 40, 67] that spatially varying post-filters can achieve similar spatial resolution and noise properties as that of regularized reconstruction. The post-filters used in this section are more restrictive since we require them to have the form of (6.4).

6.1.3 Spatial Resolution of z-dimension Post-Filtering

We use $\boldsymbol{R}_z$ to denote the z-dimension post-reconstruction regularizer, which has the frequency response

$$R_z(\varrho, \Phi, \Theta) = (2\pi\varrho \sin\Theta)^2 \tag{6.5}$$

if first-order differencing is used, and

$$R_z(\varrho, \Phi, \Theta) = (2\pi\varrho \sin\Theta)^4 \tag{6.6}$$

for second-order differencing. Combining (6.4) with the the local frequency reponse associated with 2D xy plane regularization, The local frequency response associated with 2D regularization followed by z-dimension post-reconstruction denoising is

$$\mathcal{L}_{\mathrm{2Dzpf}} = \frac{1}{1+\zeta(2\pi)^2(\varrho\cos\Theta)^3} \frac{1}{1+\zeta_z R_z(\varrho, \Phi, \Theta)}.$$

Unfortunately, this expression is quite different from the target (3.35) and the corresponding PSFs may have a much different shape than the target. In fact it appears im-possible to provide completely isotropic resolution using 2D regularization followed by z-dimension post-reconstruction denoising. Therefore we simply tune the parameters ζ and ζ_z so the FWHM of the PSF matches the target FWHM.

6.1.4 3D Post-Filter

We use R_{3Dpf} to denote the 3D post-reconstruction regularizer. The frequency response of 3D post-reconstruction denoising follows (6.4), where

$$R_{\text{3Dpf}} = (2\pi\varrho)^2 \tag{6.7}$$

for first-order differencing and

$$R_{\text{3Dpf}} = (2\pi\varrho)^4 \tag{6.8}$$

for second-order differencing. The frequency response of 3D post-reconstruction denoising is

$$\mathcal{L}_{\text{3Dpf}} = \frac{1}{1 + \zeta R_{\text{3Dpf}}}. \tag{6.9}$$

This expression is much more similar to the target, but is still different because the exponent of ϱ is either two or four depending on the differencing. To attempt a fair comparison in terms of resolution matching, we tune ζ again so that the FWHM of the PSF matches the target.

6.1.5 Plots of Point Spread Functions

Fig. 6.3 shows normalized profiles cut along the X, Y, and Z axis of PSFs associated with these three methods with the regularization parameters tuned such that the FWHM$\approx$ 1.43 pixels. The PSFs associated with these three methods are very similar. This is consistent with Stayman's results [67] that post-filtering can achieve similar spatial resolution properties as that of quadratically regularized reconstruction, even with the structural restriction made in (6.4).

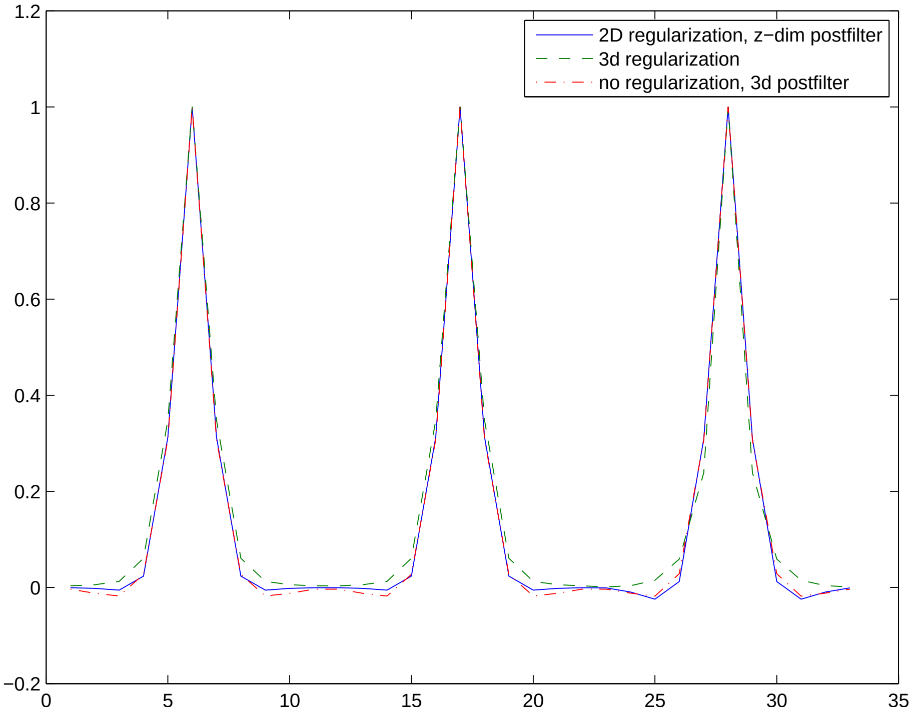


Figure 6.3: Profiles cut through PSFs in different directions concatenated into one array. From left to right, along the X-axis, Y-axis, and Z-axis.

Next we plot several PSFs using these three methods using statistical weighting $\boldsymbol{W}$ from a simulated sinogram from the NCAT [57] phantom. Fig. 6.4 displays slices of the PSF associated with 3D post-reconstruction denoising and contour plots of those slices. We only plot this at one location since it is spatially invariant. Fig. 6.5-6.8 display slices of PSFs for 3D regularization and 2D regularization with z-dimension post-reconstruction denoising at several locations through the volume.

As is quite apparent, 2D regularization with z-dimension post-reconstruction denoising provides for more isotropy than 3D regularization under realistic weightings. One may notice extremely streak like PSFs at several locations. In designing these regularizers, we used the *AIMA* method with $\alpha = 0$. The streaks are caused by too many r_l^j values being 0, which can be fixed by raising α, as discussed in §4.1.3. We choose to use $\alpha = 0$ in this section because α controls the tradeoff between *AIMA* regularization and certainty based

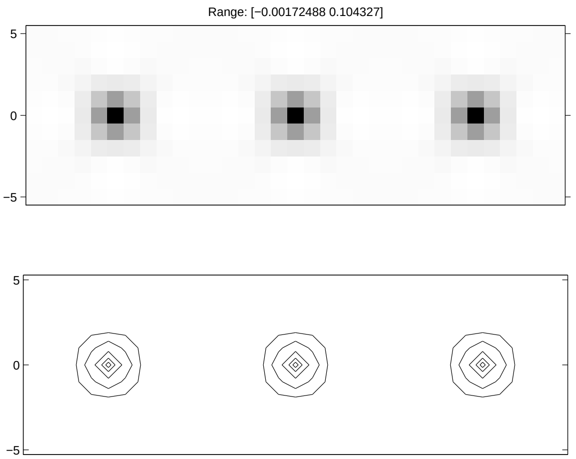


Figure 6.4: from left to right, xy, xz, yz slices of the PSF from 3D post-reconstruction denoising.

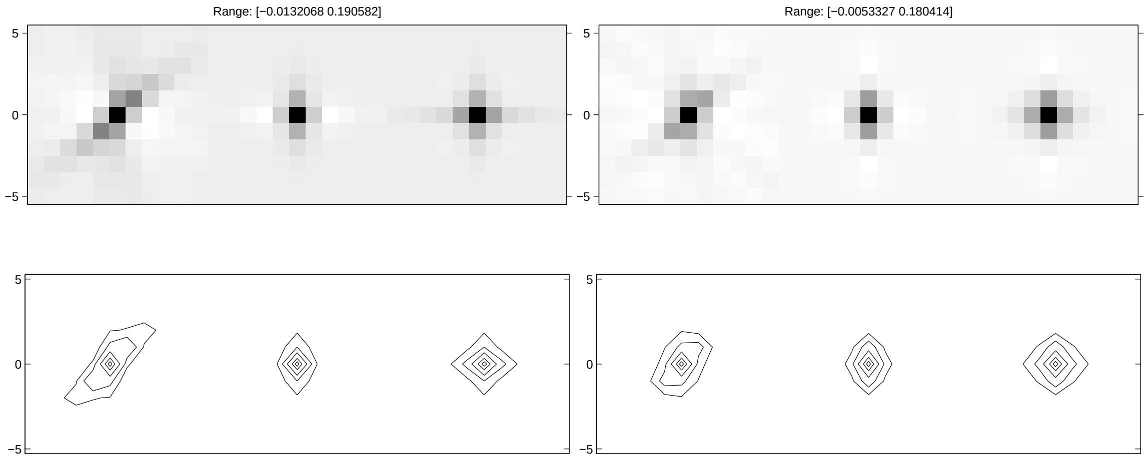


Figure 6.5: PSF at (100,-80,0), using 3D regularization on the left, using 2D regularization with a z-dimension post-reconstruction denoising on the right. Each plot displays, from left to right, xy, xz, yz slices of the PSF.

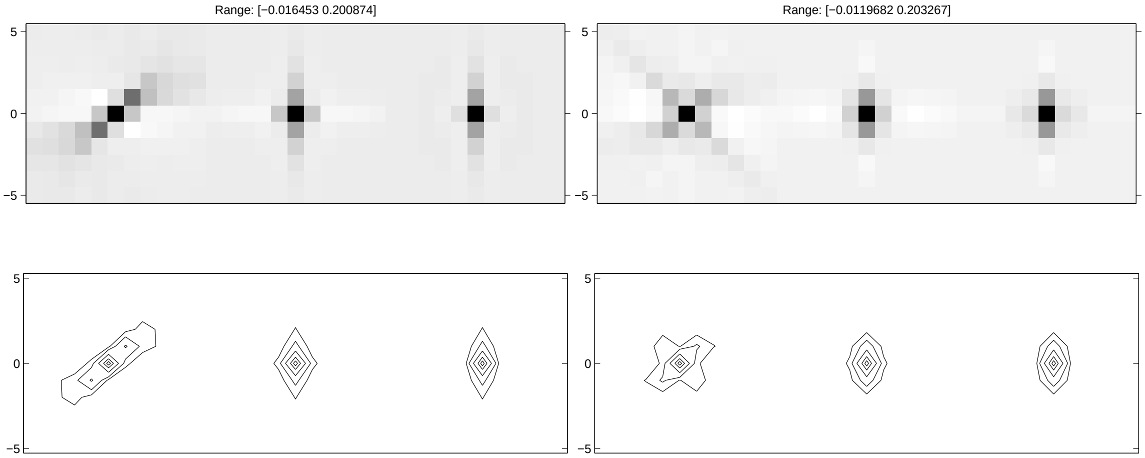


Figure 6.6: PSF at (-100,-100,0), using 3D regularization on the left, using 2D regularization, with z-dimension post-reconstruction denoising on the right. Each plot displays, from left to right, xy, xz, yz slices of the PSF.

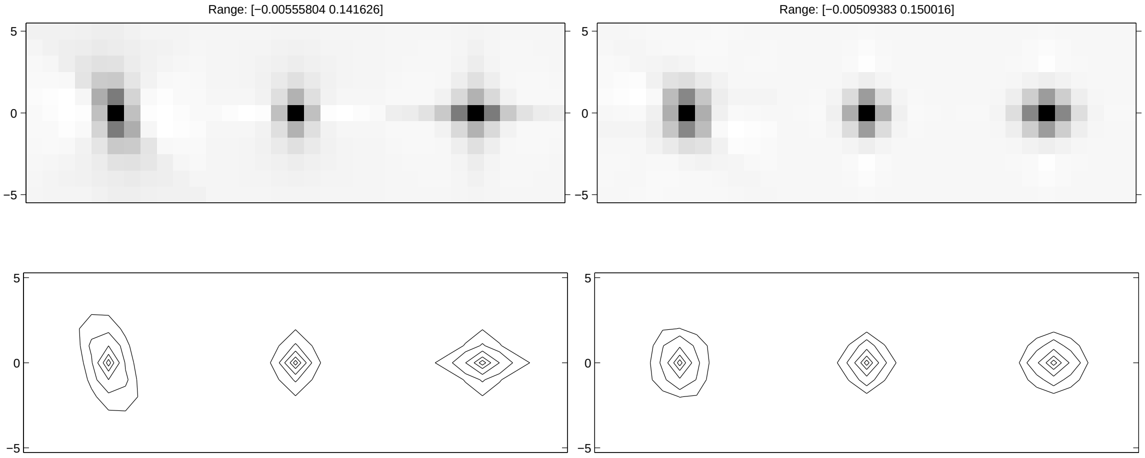


Figure 6.7: PSF at (100,0,-40), using 3D regularization on the left, using 2D regularization with a z-dimension post-reconstruction denoising on the right. Each plot displays, from left to right, xy, xz, yz slices of the PSF.

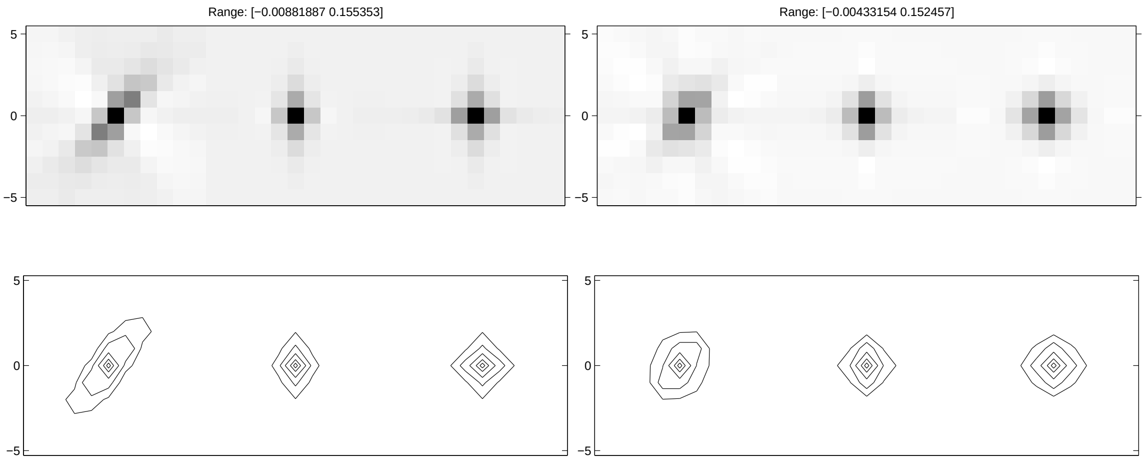


Figure 6.8: PSF at (-100,0,20), using 3D regularization on the left, using 2D regularization with a z-dimension post-reconstruction denoising on the right. Each plot displays, from left to right, xy, xz, yz slices of the PSF.

regularization and we wish to give the reader a sense of pure *AIMA* regularization. Next, we will look at the variance of these three methods to gain some insight into the noise properties.

6.2 Noise Properties

Analyzing the covariance for non-quadratic edge preserving regularization is extremely difficult, though there have been various attempts and approximations for doing so [3, 42], However in the quadratic case using a PWLS estimator we can develop exact expressions for the variance [16]. The covariance for a PWLS estimator minimizing the following cost function

$$\hat{\boldsymbol{x}}_{\text{PWLS}} = \arg\min_{\boldsymbol{x} \geq 0} \frac{1}{2} ||\boldsymbol{A}'\boldsymbol{x} - \ell(\boldsymbol{y})||^2_{\boldsymbol{W}} - \zeta R(\boldsymbol{x}), \tag{6.10}$$

is

$$\mathsf{Cov}\{\hat{\boldsymbol{x}}_{\text{PWLS}}\} = [\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A} + \zeta \boldsymbol{R}]^{-1} \boldsymbol{A}'\boldsymbol{W} \, \mathsf{Cov}\{\ell(\boldsymbol{y})\} \, \boldsymbol{W}\boldsymbol{A}[\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A} + \zeta \boldsymbol{R}]^{-1}, \tag{6.11}$$

where $\boldsymbol{R}$ is the Hessian of the regularizer, $R(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}'\boldsymbol{R}\boldsymbol{x}$. $\boldsymbol{W}$ is the user defined weighting matrix, which we have chosen in this thesis to be $w_i = \mathsf{Cov}^{-1}\{\ell_i\} \approx \mathsf{diag}\{y_i\}$ for

CT. This expression is exact if a PWLS estimator is used; a PL estimator would yield approximately the same covariance [16], Substituting in the weighting yields

$$\mathsf{Cov}\{\hat{\boldsymbol{x}}_{\mathrm{PWLS}}\} = [\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A} + \zeta\boldsymbol{R}]^{-1}\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A}[\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A} + \zeta\boldsymbol{R}]^{-1}. \tag{6.12}$$

Adding in some form of post-reconstruction denoising, $\hat{\boldsymbol{x}}_{\mathrm{post}} = \boldsymbol{L}_{\mathrm{post}}\hat{\boldsymbol{x}}_{\mathrm{PWLS}}$, where $\boldsymbol{L}_{\mathrm{post}}$ is a circulant matrix that performs a convolution with the impulse response associated with the post-reconstruction denoising, (6.12) becomes

$$\mathsf{Cov}\{\hat{\boldsymbol{x}}_{\mathrm{post}}\} = \boldsymbol{L}_{\mathrm{post}}[\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A} + \zeta\boldsymbol{R}]^{-1}\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A}[\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A} + \zeta\boldsymbol{R}]^{-1}\boldsymbol{L}'_{\mathrm{post}}. \tag{6.13}$$

The inverses would take a long time to compute because these matrices are large, however, we can compute the variance of the estimator efficiently by adding $\boldsymbol{\delta}^j$ to compute the variance at once pixel and using the circulant approximation from §III,

$$\mathsf{Var}\left\{\hat{\boldsymbol{x}^j}\right\} = \boldsymbol{\delta}^{j\prime}\boldsymbol{L}_{\mathrm{post}}[\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A} + \zeta\boldsymbol{R}]^{-1}\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A}[\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A} + \zeta\boldsymbol{R}]^{-1}\boldsymbol{L}'_{\mathrm{post}}\boldsymbol{\delta}^j. \tag{6.14}$$

Factoring out a Fourier Matrix $\boldsymbol{Q}$ which is centered at voxel j,

$$\begin{aligned}\mathsf{Var}\left\{\hat{\boldsymbol{x}^j}\right\} &= \boldsymbol{\delta}^{j\prime}\boldsymbol{Q}'\left[\frac{\boldsymbol{\Lambda}^j(\boldsymbol{\Upsilon}^j)^2}{(\boldsymbol{\Lambda}^j + \zeta\boldsymbol{\Gamma}^j)^2}\right]\boldsymbol{Q}\boldsymbol{\delta}^j \\ &= \sum_k \frac{\boldsymbol{\Lambda}^j_k|\boldsymbol{\Upsilon}^j_k|^2}{(\boldsymbol{\Lambda}^j_k + \zeta\boldsymbol{\Gamma}^j_k)^2},\end{aligned} \tag{6.15}$$

where as in §III, $\boldsymbol{\Lambda}^j$ is the DFT of $\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A}\boldsymbol{\delta}^j$, $\boldsymbol{\Gamma}^j$ is the DFT of $\boldsymbol{R}\boldsymbol{\delta}^j$, and $\boldsymbol{\Upsilon}^j$ is the DFT of $\boldsymbol{L}_{\mathrm{post}}\boldsymbol{\delta}^j$.

We used (6.15) to compute the standard deviation of 2D regularization with z-dimension post-reconstruction denoising, 3D regularization, and 3D post-reconstruction denoising for different values of ζ. Then, we plotted the variances as a function of the spatial resolution associated with ζ in Fig. 6.9. Since $\boldsymbol{W}$ is object dependent and it is impractical for us to calculate standard deviation plots for all possible values of $\boldsymbol{W}$ we have assumed

$\boldsymbol{W} = \boldsymbol{I}$ in these plots, which is equivalent to assuming the noise is additive white Gaussian noise. While this is an unrealistic Assumption to make for tomographic systems, these calculations should still give us an indication of the performance of these three methods even when $\boldsymbol{W} \neq \boldsymbol{I}$. In the derivation of certainty based regularization in [23], Fessler made the following approximations,

$$\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A} \approx \boldsymbol{D}\boldsymbol{A}'\boldsymbol{A}\boldsymbol{D}$$

which motivated a regularization design of the form

$$\boldsymbol{R} = \boldsymbol{D}\boldsymbol{R}_0\boldsymbol{D},$$

where $\boldsymbol{D} = \mathsf{diag}\{\kappa\}$. Using these equations, (6.14) becomes

$$\mathsf{Var}\{\hat{\boldsymbol{x}}^j_{\text{post}}\} \approx \boldsymbol{\delta}^{j'}\boldsymbol{D}^{-1}\boldsymbol{L}'_{\text{post}}[\boldsymbol{A}'\boldsymbol{A} + \zeta\boldsymbol{R}_0]^{-1}\boldsymbol{A}'\boldsymbol{A}[\boldsymbol{A}'\boldsymbol{A} + \zeta\boldsymbol{R}_0]^{-1}\boldsymbol{L}_{\text{post}}\boldsymbol{D}^{-1}\boldsymbol{\delta}^j.$$

When $\boldsymbol{W} \neq \boldsymbol{I}$, the variances calculated in Fig. 6.9 are subject to some scale factor $\boldsymbol{D}^{-1}$ but the overall relationships remain the same. The standard deviations are similar for all three estimators. The noise properties appear similar in the quadratic case. There are differences however because matching the FWHM of profiles through the PSF still leaves room for differences in spatial resolution.

6.3 Reconstructed Images

Stayman argued that for quadratic regularization, the same spatial resolution and noise properties could be achieved with the appropriate post-filter. However the non-quadratic case is much more difficult to analyze. Anecdotal results suggest that applying edge preserving post-reconstruction denoising to unregularized reconstructions may accentuate streak artifacts that are common to unregularized tomographic images. Below is a 2D example, Fig. 6.10 has been reconstructed using certainty based regularization with a Huber penalty, and Fig. 6.11 was reconstructed with unregularized WLS and then we applied

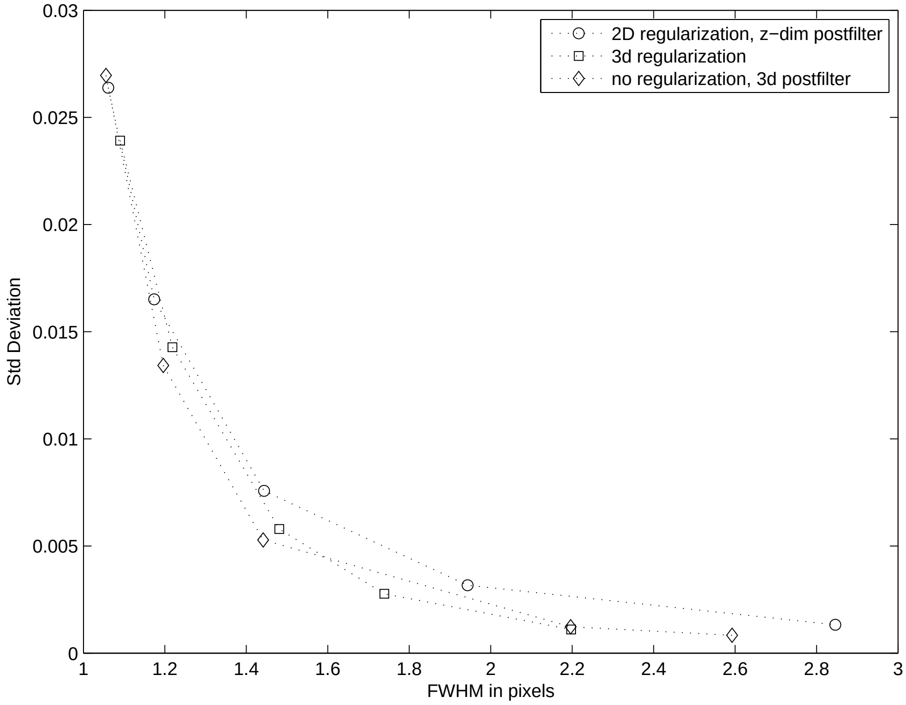


Figure 6.9: standard deviations of the 3 methods.

an edge preserving post-reconstruction denoising of the form (6.3) using the same Huber penalty with matched FWHM. More streak artifacts are present in the post-filtered image than in the regularized image.

While these results suggest that post-reconstruction edge-preserving denoising applied to unregularized reconstructions is suboptimal, we believe that since the noise of 2.5D systems is statistically independent in the z-direction, edge-preserving z-dimension post-reconstruction denoising will not have the same problems as edge-preserving post-reconstruction denoising applied to the XY plane. These streaks exist because the denoising of (6.3) uses an unweighted norm which assumes that the noise is white. One could choose $\boldsymbol{W} = \mathsf{Cov}^{-1}\{\hat{\boldsymbol{x}}_{\mathrm{PWLS}}\}$, however computing that quantity is extremely slow and suboptimal because of the size of the matrices involved.

We next simulate a 2.5D system using a stack of 2D systems to form a 51 slice system with a similar geometry to [70] with $N_x = N_y = 256$ and $N_z = 51$, and $\Delta_x = \Delta_y =$

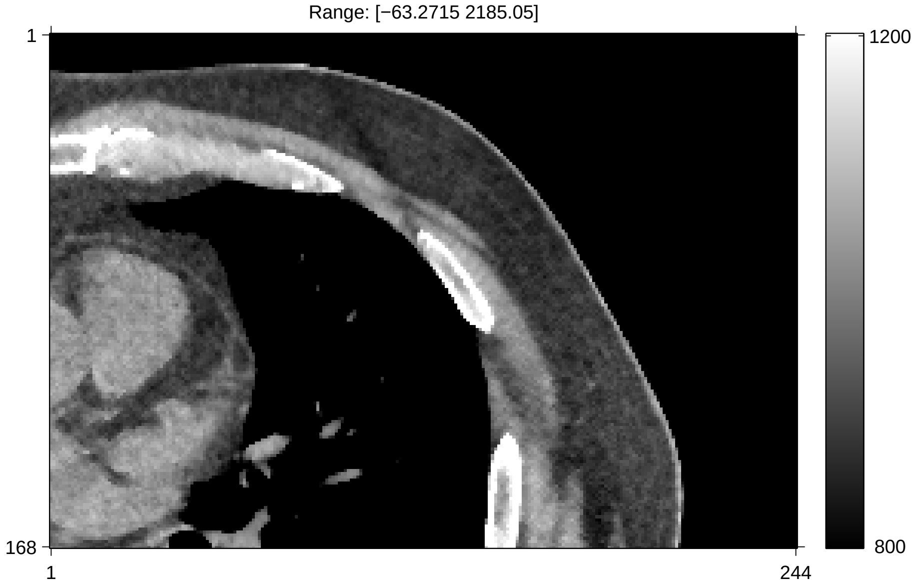


Figure 6.10: Reconstruction using regularization.

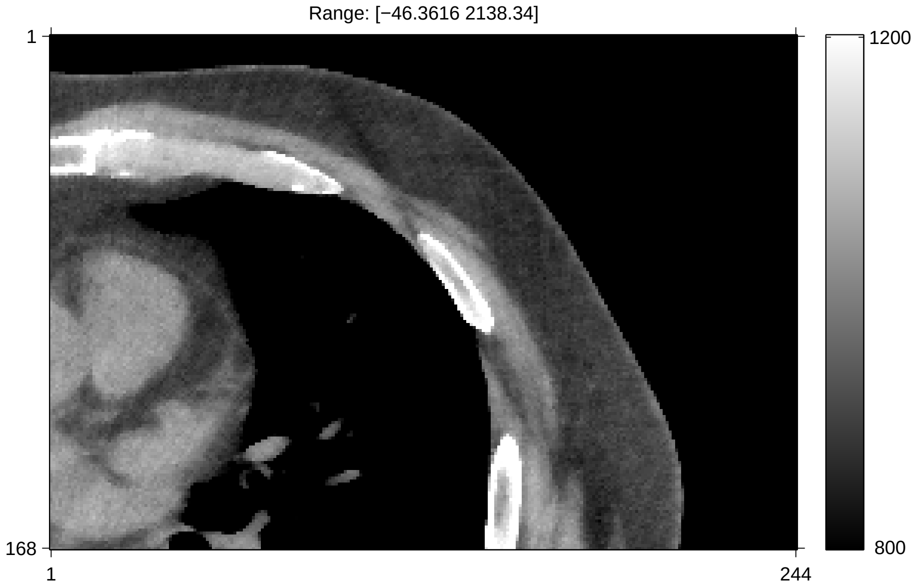


Figure 6.11: Reconstructed using post-reconstruction denoising.

$\Delta_z = 1.95mm$. To simulate model-mismatch, we used a $1024 \times 1024 \times 51$ NCAT phantom and generated a sinogram using the same distance-driven backprojector except with $N_x = N_y = 1024$ and $N_z = 51$, $\Delta_x = \Delta_y = 0.49mm$ and $\Delta_z = 1.95mm$. We compared 4 reconstructions using different methods: (i) 3D regularization designed using the *AIMA* method with $\alpha = 0.1$ and 5 neighbors, the four associated with 2D regularization and an additional neighbor in the z direction. (ii) 2D regularization designed using the *AIMA* method with $\alpha = 0.1$ and 4 neighbors, followed by z-dimension post-reconstruction edge-preserving denoising. (iii) extremely small certainty based regularization (just to help the algorithm converge) and then 3D post-reconstruction edge-preserving denoising. (iv) 3D regularization using certainty based weightings [23]. Fig. 6.12-6.17 display axial images for slices 26 and 48, and coronal images for slices 129 and 100 for the above 4 methods as well as the phantom, and the minimally regularized (almost no-denoising) reconstruction. The images were reconstructed using an ordered subsets algorithm with separable paraboloidal surrogates [63, 64] with 10 iterations using 246 subsets, 20 iterations using 82 subsets, 40 iterations using 41 subsets, 100 iterations with 2 subsets, and 200 iterations with 1 subset. To ensure convergence of the minimally regularized reconstruction which converges slower than other regularized reconstructions, we performed an additional 800 iterations with 1 subset.

As mentioned previously, Fig. 6.13 was reconstructed using slight regularization to help the algorithm converge. This gives us a sense of the noise that is present in unregularized reconstructions. The coronal images are all quite similar, which confirms our previous assumption that since under 2.5D geometries, noise in different slices are statistically independent and thus all of our regularization techniques should have similar performance in the z dimension. The axial images are quite different. We can see from Fig. 6.13 that there is a substantial amount of streak-like noise. The reconstructions

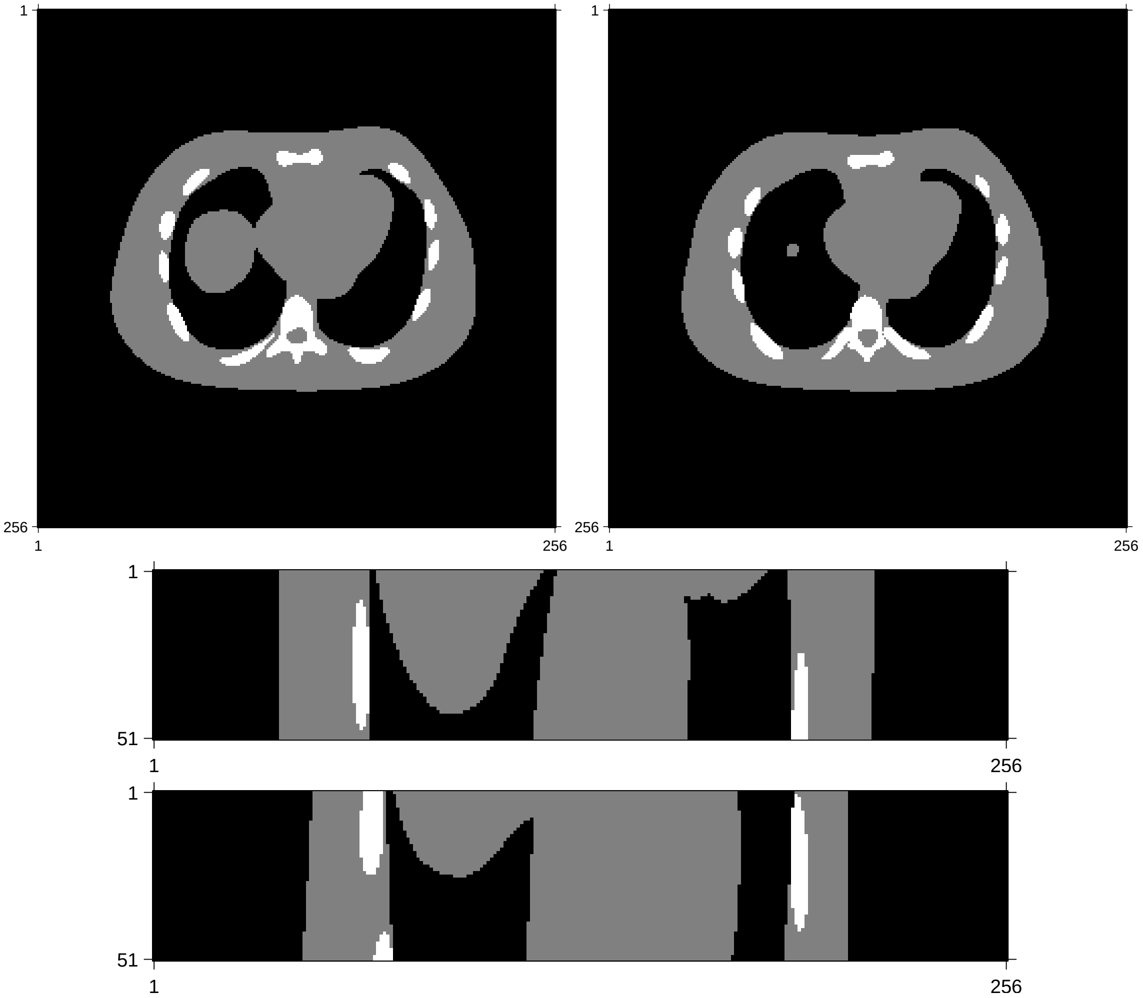


Figure 6.12: Axial and Coronal images of the phantom.

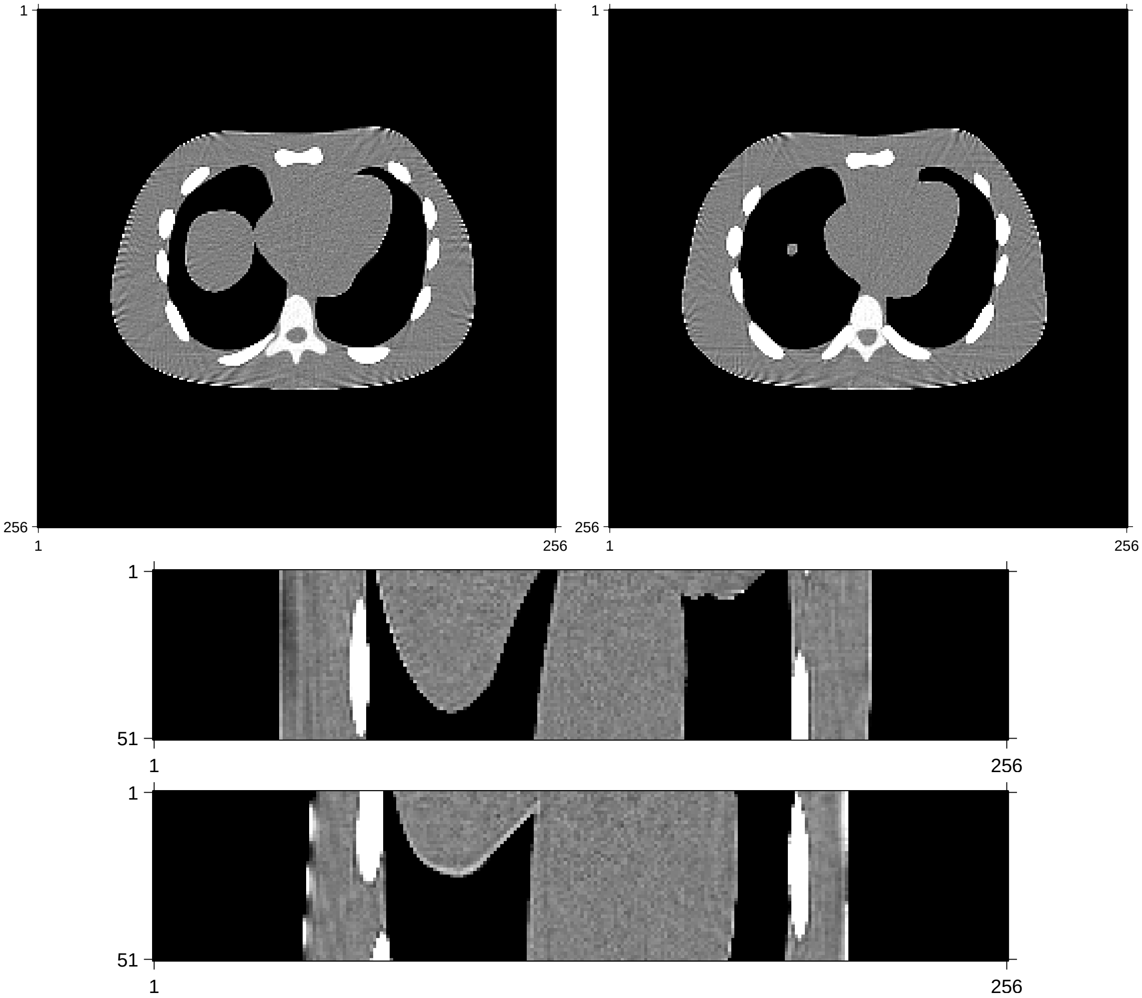


Figure 6.13: Axial and Coronal images reconstructed using minimal regularization.

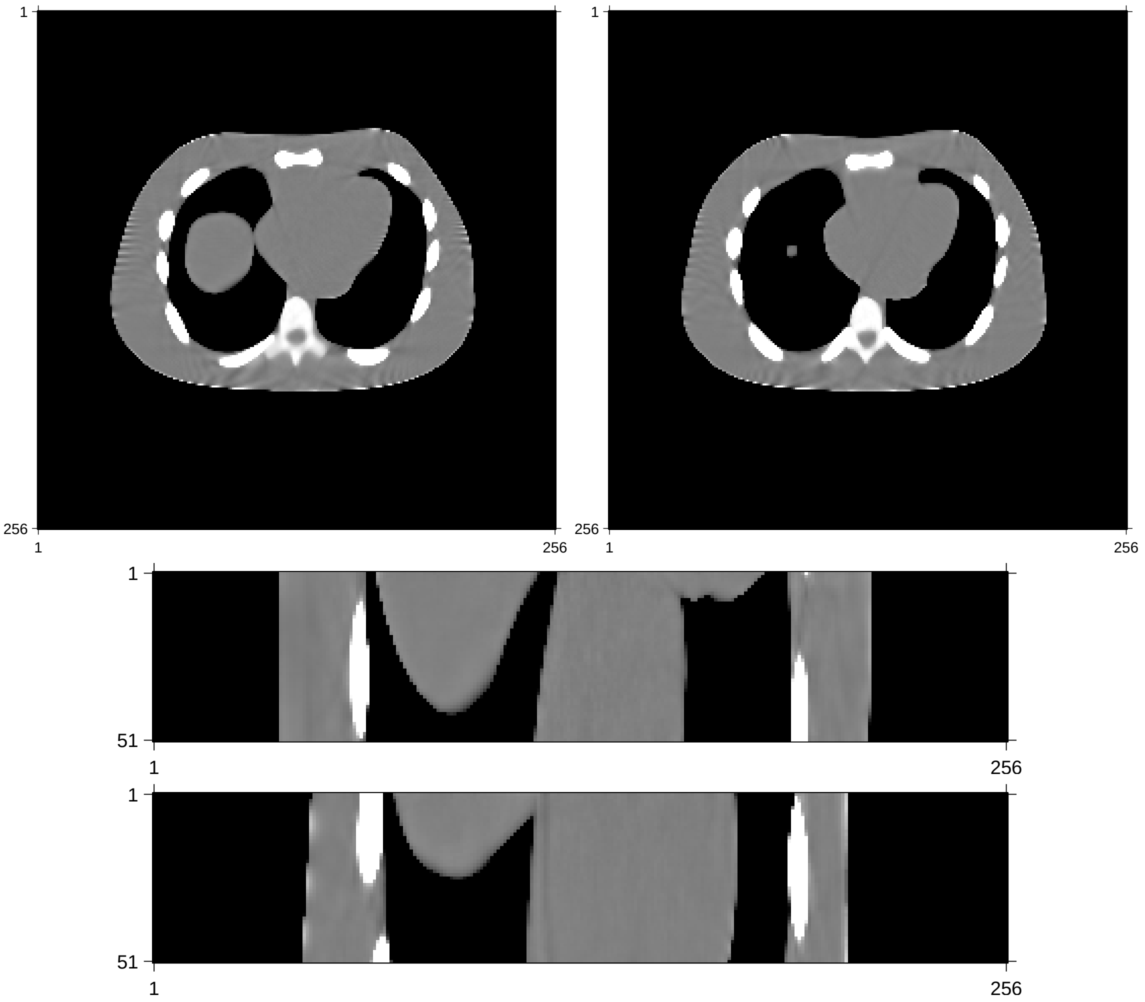


Figure 6.14: Axial and Coronal images reconstructed using 3D edge-preserving regularization.

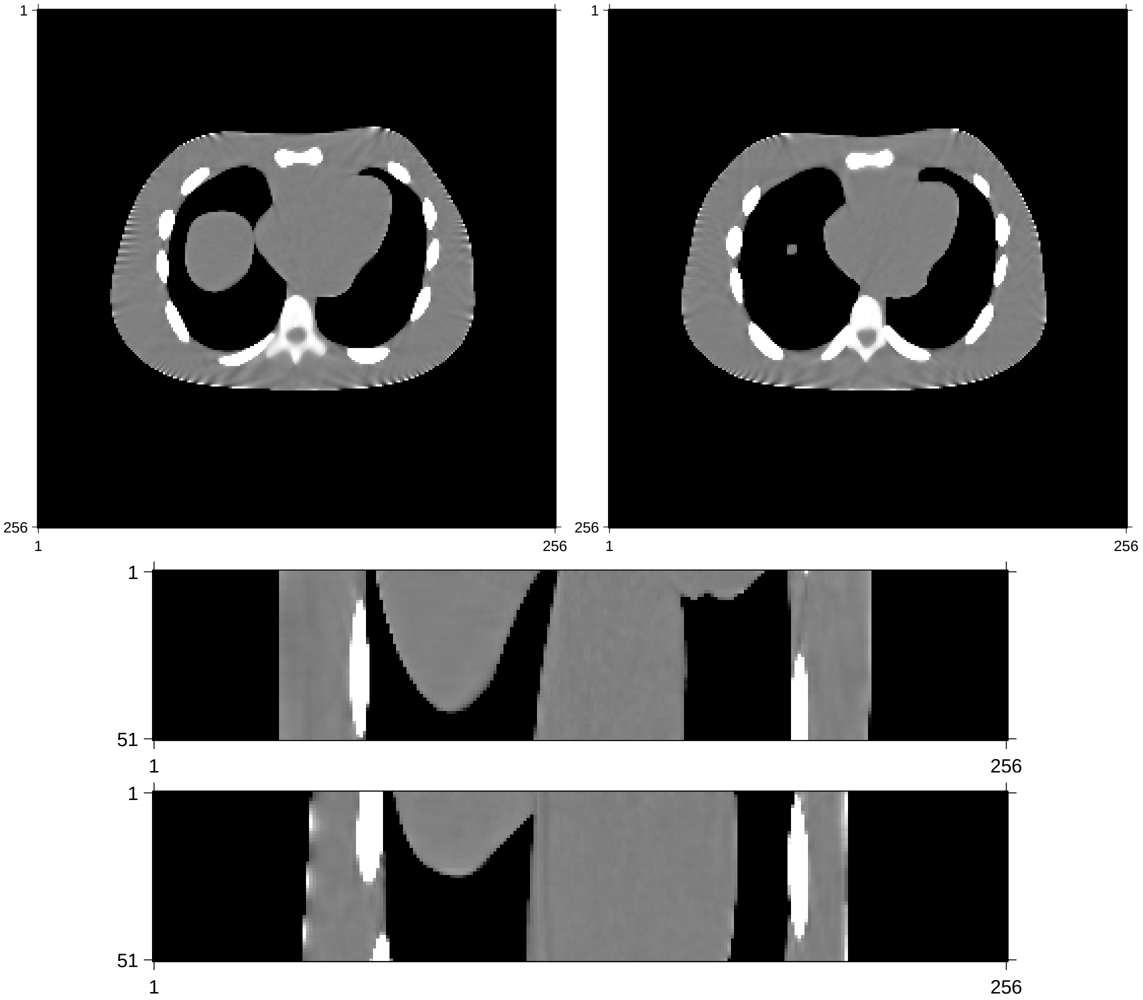


Figure 6.15: Axial and Coronal images reconstructed using 2D edge-preserving regularization and z-dimension edge-preserving denoising.

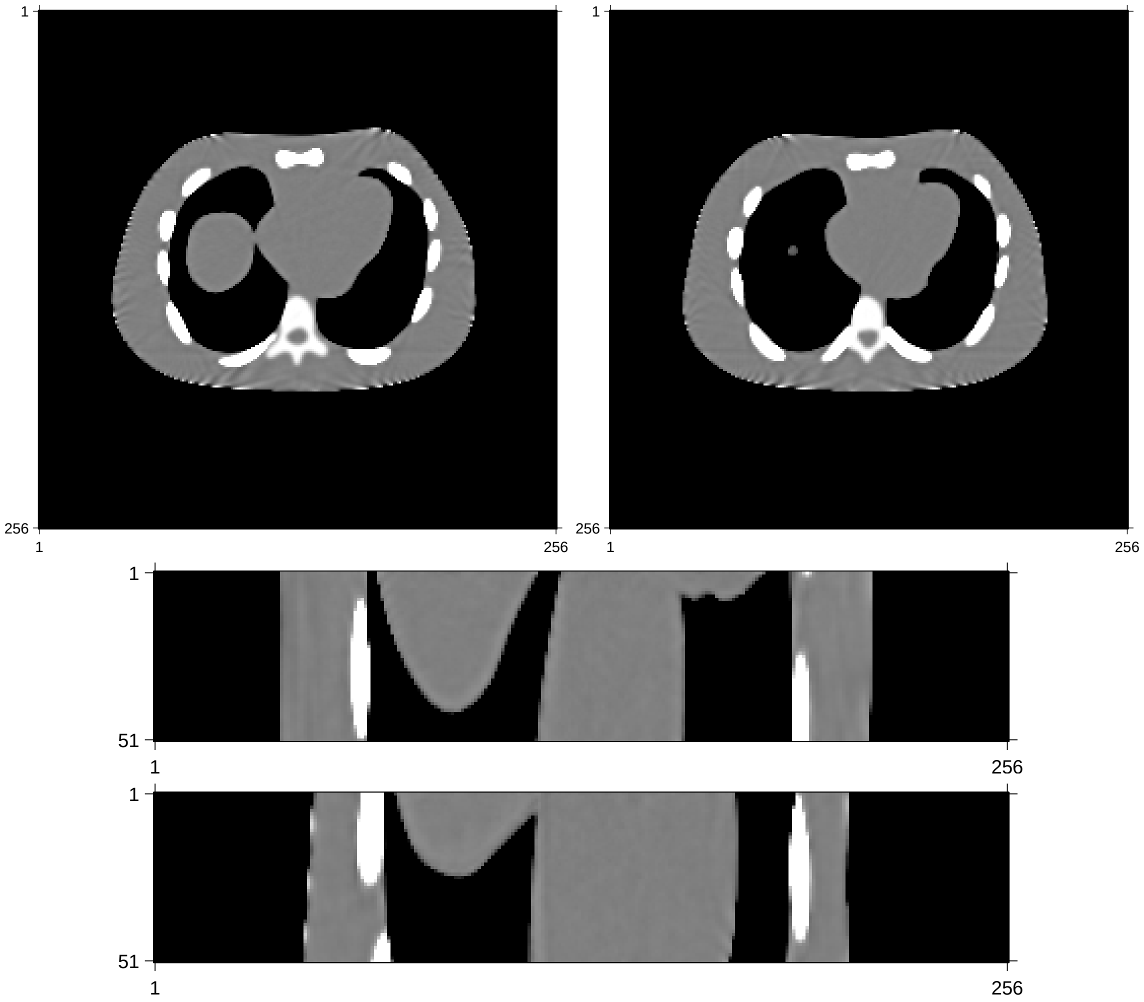


Figure 6.16: Axial and Coronal images reconstructed using minimal regularization and then 3D iterative edge-preserving denoising.

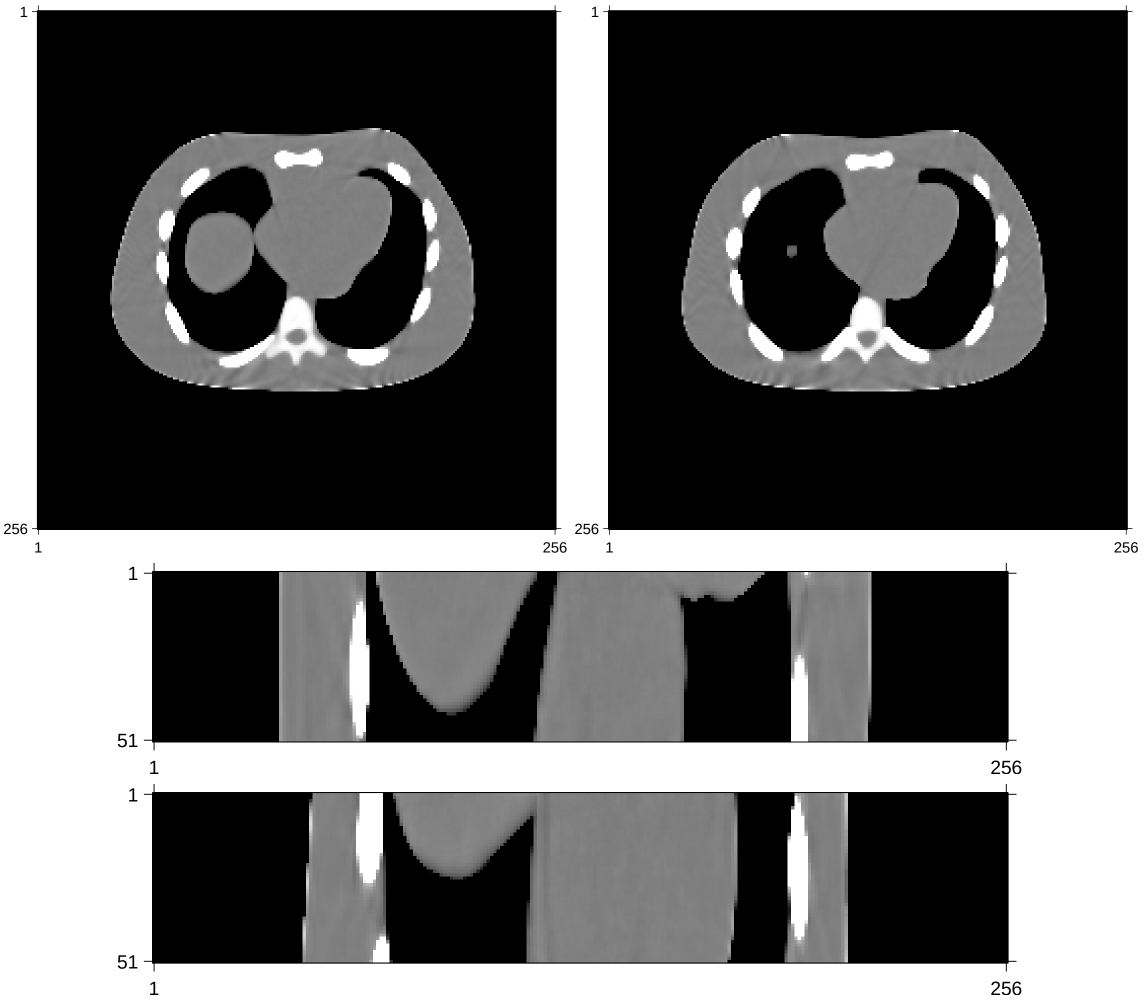


Figure 6.17: Axial and Coronal images with certainty based edge-preserving regularization.

from 3D post-reconstruction edge-preserving regularization seems to preserve many of these streaks. The reconstructions using 3D and 2D *AIMA* regularization in Fig. 6.15 and Fig. 6.14 appear to have even more streaks than Fig. 6.13. Reconstructions using certainty based regularization in Fig. 6.17 have the fewest streaks and suggest that edge preserving regularization does have the potential to reduce streaks. The obvious question is to ask why edge-preserving *AIMA* regularization has more streaks than edge-preserving certainty based regularization. There are several possibilities. First, Ahn suggested in [3] that certainty based weights designed for quadratic regularization could be extended to edge-preserving regularization. There may be issues generalizing that work to the parameterization of the regularization we are using in this thesis. Second, we have experimented with the parameter α that controls the tradeoff between *AIMA* regularization and certainty based regularization in an attempt to eliminate zeros in the Hessian. However, this is a crude method of quality control. Regularization designed using *AIMA* gives us the best fit to a desired target frequency response, however there is no guarantee on the shape of the associated PSF. Developing more sophisticated ways of controlling the quality of the *AIMA* approach is an open problem.

6.4 Summary

In this chapter, we have considered 3 approaches to denoising and image reconstruction: (i) 3D edge-preserving regularization, (ii) 2D edge-preserving regularization with z-dimension post-reconstruction edge-preserving denoising, and (iii) 3D post-reconstruction edge-preserving denoising. In the quadratic case, using an NCAT phantom, method (i) provides for completely isotropic and uniform spatial resolution. Method (ii) provides for improved isotropy and uniformity over (iii). We hypothesized that method (i) would be undesirable because unregularized reconstructions converge very slowly. We also hypothesized that post-reconstruction edge-preserving denoising would preserve streak artifacts present in unregularized reconstructions. We believed that method (ii) would be able to achieve good isotropic spatial resolution, and eliminate streaks that occur mostly in the xy plane. Simulation results demonstrate that while in the quadratic case, the 3 methods have similar spatial resolution and noise properties, in the non-quadratic edge-preserving case, they are quite different. Methods (ii) and (iii) appear to preserve even more streaks than method (i). We tried using edge-preserving certainty based 3D regularization which seemed to eliminate the most streaks. While the *AIMA* approach to regularization design used in methods (ii) and (iii) pick the best regularization coefficients given our constraints, there is no quality control over the properties of those regularizers. We believe that using the parameter α which controls the tradeoff between *AIMA* regularizers and certainty-based regularizers will allow the user to construct regularizers that have desirable spatial resolution and noise properties, even in the non-quadratic case. Selecting the parameter α, and other methods of quality control for *AIMA* regularization is an open problem.

CHAPTER VII

Conclusions and Future Work

7.1 Summary

In this dissertation, we have proposed quadratic regularization design techniques for parallel and fan-beam geometries for 2D and 3D systems. These methods have varying degrees of performance in terms of spatial resolution, noise properties, and computational performance.

For 2D, we have extended a parallel-beam regularization design technique based on continuous space analogs, the *AIMA* method, to fan-beam geometries. This method extracts a continuous weighting function from a user-defined weighting matrix using the fan-beam geometry and the corresponding Jacobian. This technique is computationally efficient and can be solved mostly analytically, and we have shown that it performs well in terms of resolution isotropy.

For 3D, we have extended 2D regularization design approaches to 3D systems. This method is slower, as 3D volumes are larger and there does not appear to be a corresponding analytical solution for the 3D case. 3D regularization performs well for simple phantoms, but breaks down with more realistic data because it is trying to counteract a fundamental anisotropy in 3D systems with small cone or cross-plane angles. This regularization design method also requires larger amounts of memory than the 2D case, because it needs to store more coefficients for more voxels.

As an alternative to the *AIMA* method, we have developed a slower, but more accurate *FIIN* method. This method avoids using the *AIMA* approximation which makes the *AIMA* method faster. While this method performs better in terms of resolution isotropy, the benefits are not significant. This method is a suitable alternative to the *AIMA* method when resolution isotropy is a priority over computational efficiency.

These regularization design methods have the potential to yield too few non-zero coefficients due to the non-negativity constraint. In this dissertation, we have developed a method for guarding against this by controlling a tradeoff between the *AIMA* and *FIIN* methods and the certainty based regularization design approach.

As an alternative to 2D and 3D regularization, we have investigated 2D regularization followed by z-dimension post-reconstruction denoising. This combined approach benefits from the improved spatial resolution and computational performance of the *AIMA* method for 2D systems, and achieves excellent overall isotropy over 3D regularization. We have analyzed the spatial resolution and noise properties of this method in the quadratic case, and found it to be comparable to 3D regularization. We have also performed simulations using this method with non-quadratic edge-preserving regularization, and found that greater work needs to be done to control the quality of reconstructed images.

7.2 Future Work

- This work focused on using a 2D derivation for the Grammian operator which we have found to be suitable for 3D systems with small cross-plane or cone angles. However, it would be desirable to extend the derivation of the Grammian operator to 3D geometries, and explore 3D regularization design with the 3D Grammian.

- This work focused on 2.5D geometries. Extending this work to fully 3D PET and cone-beam CT can be done using the 3D Grammian, or by other methods.

- Extensions to second order differencing would give us more basis functions and more degrees of freedom which will improve the accuracy of our regularization design.
- This work focused on equations for spatial resolution, and noise for quadratic regularization. However, in CT, non-quadratic edge-preserving regularization is the preferred method for de-noising. It would be desirable to extend our analysis to edge-preserving regularization, and analyze the local-perturbation response [3]. This would help us understand more deeply the trade-off between regularization and post-filtering.
- We found no apparent analytical solution to 3D regularization design. It is possible that one does not exist, however further exploration may lead to an efficient algorithm for 3D regularization design.
- The regularization design methods used in this dissertation are overly rigorous, we calculate coefficients for every pixel of voxel using every projection angle in the sinogram. A coarser scheme for regularization design, and angular sub-sampling would help make regularization design methods, especially the 3D methods, more efficient. A coarser system for regularization design may also use less memory, and make 3D regularizers more practical. The performance tradeoffs for such a scheme must be explored.
- The 2.5D systems considered in this thesis are unrealistic. More work needs to be done to develop regularization for fully 3D systems, or to determine the efficacy of this regularization design with fully 3D systems.
- The proposed regularization design can lead to strangely shaped PSFs that can cause artifacts in the resulting image. We had proposed a user defined parameter α which

could control a tradeoff between *AIMA* regularization and certainty based regularization which does not have this issue. More work needs to be done to determine how best to choose this parameter. Other ways to ensure that the spatial resolution properties of designed regularizers are desirable need to be investigated.

- In trying to understand the causes of streaks in edge-preserved regularization, we analyzed the quadratic case. With the NCAT phantom we saw streaks surrounding bones that we had not seen before. Further study is required to understand the causes of these streaks and why they are worse in quadratically regularized images than the edge-preserving case.

APPENDICES

APPENDIX A

Fourier Analysis of Fan Beam Geometries

This appendix considers fan-beam geometries and uses continuous-space analysis to analyze the Fourier transform of the Grammian operator $\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A}$ to simplify the regularization design problem in (3.8). One can use polar coordinates Φ, ρ, and continuous-space analysis to separate the angular and radial components of $\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A}$. Using this framework, isotropy can be thought of as eliminating dependence on the angular component Φ.

A.1 Fan-Beam Geometry

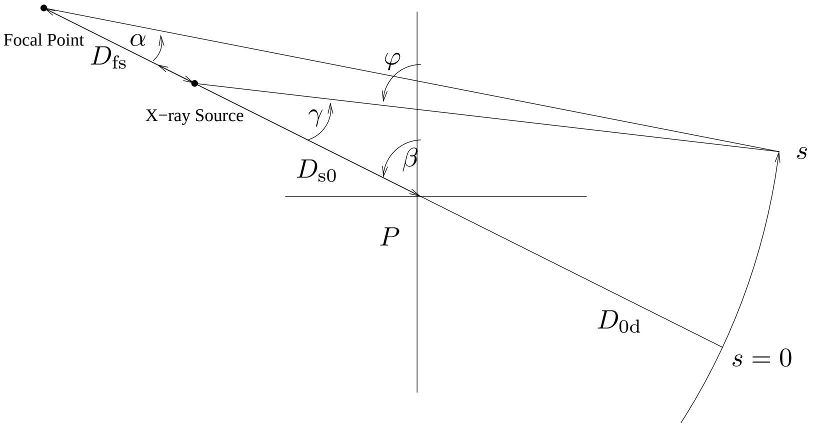


Figure A.1: Illustration of Fan beam geometry.

Fig. A.1 illustrates the fan-beam geometry that we consider. Let P be the rotation isocenter. D_{0d} denotes the distance from the point P to the detector, D_{s0} denotes the distance from the X-ray source to P, and D_{fs} denotes the distance from the X-ray source to the focal point of the detector arc. Define $D_{\text{sd}} \triangleq D_{0\text{d}} + D_{\text{s}0}$ to be the total distance from the X-ray source to the center of the detector, and $D_{\text{fd}} \triangleq D_{\text{sd}} + D_{\text{fs}}$ to be the total distance from the focal point to the center of the detector. This formulation encompasses a variety of system configurations by allowing the detector focal point to differ from the X-ray source location. For flat detectors, $D_{\text{fs}} = \infty$. For third-generation X-ray CT systems, $D_{\text{fs}} = 0$. For fourth generation X-ray CT systems, $D_{\text{fs}} = -D_{\text{s}0}$.

Let $s \in [-s_{\max}, s_{\max}]$ denote the (signed) *arc length* along the detector, where $s = 0$ corresponds to the detector center. Assuming detector elements are equally spaced along the detector, arc length is the natural parameterization. The various angles have the following relationships:

$$\begin{aligned} \alpha(s) &= \frac{s}{D_{\text{fd}}} \\ \gamma(s) &= \tan^{-1}\left(\frac{(D_{\text{fd}})\sin\alpha(s)}{(D_{\text{fd}})\cos\alpha(s) - D_{\text{fs}}}\right) \end{aligned}$$

where the two most important cases are

$$\gamma(s) = \begin{cases} \alpha(s), & D_{\text{fs}} = 0 \\ \tan^{-1} s/D_{\text{sd}}, & D_{\text{fs}} = \infty. \end{cases} \tag{A.1}$$

The relationship between s and γ is:

$$s = \begin{cases} (D_{\text{fd}}) \left[\gamma - \arcsin\left(\frac{D_{\text{fs}}}{D_{\text{fd}}} \sin\gamma\right)\right], & 0 \le D_{\text{fs}} < \infty \\ D_{\text{sd}} \tan\gamma, \quad D_{\text{fs}} = \infty. \end{cases} \tag{A.2}$$

The ray corresponding to detector element s and angle β is

$$\mathcal{L}(s, \beta) = \left\{(x, y) : x \cos\varphi(s, \beta) + y \sin\varphi(s, \beta) = r(s)\right\},$$

where

$$\begin{aligned} \varphi(s, \beta) &\triangleq \beta + \gamma(s) \\ r(s) &\triangleq D_{\text{s}0} \sin\gamma(s)\,. \end{aligned} \tag{A.3}$$

The range of r depends on the position of the X-ray source and the size of the detector:

$$|r(s)| \leq r_{\max} \triangleq D_{s0} \sin \gamma_{\max}, \tag{A.4}$$

where $\gamma_{\max} \triangleq \gamma(s_{\max})$ and $s_{\max}$ is half the total detector arc length. The radius $r_{\max}$ defines the circular *field of view* of the imaging system. The *fan angle* is $2\,\gamma_{\max}$.

The line-integral projection $p(s, \beta)$ of f along $\mathcal{L}(s, \beta)$ is[1]:

$$\begin{aligned} p(s, \beta) &= \int_{\mathcal{L}(s,\beta)} f(x, y) \, \mathrm{d}\ell \\ &= \iint f(x, y) \\ & \delta(x \cos(\varphi(s, \beta)) + y \sin \varphi(s, \beta) - r(s)) \, \mathrm{d}x \, \mathrm{d}y, \end{aligned} \tag{A.5}$$

for $|s| \leq s_{\max}$ and $0 \leq \beta < \beta_{\max}$. We require $\beta_{\max} \geq \pi + 2\,\gamma_{\max}$ to ensure complete sampling of the FOV (otherwise the impulse response would be highly anisotropic) .

The usual inner product for fan-beam projection space is

$$\langle p_1,\, p_2 \rangle = \int_{-s_{\max}}^{s_{\max}} \int_0^{\beta_{\max}} p_1(s, \beta)\, p_2(s, \beta) \, \mathrm{d}s \, \mathrm{d}\beta \,.$$

Using this inner product in projection space, and the usual $\mathcal{L}^2$ inner product in image space, the adjoint of $\boldsymbol{\mathcal{P}}$ is given by

$$\begin{aligned} (\boldsymbol{\mathcal{P}}^* p)(x, y) &= \int_{-s_{\max}}^{s_{\max}} \int_0^{\beta_{\max}} p(s, \beta) \\ & \delta(x \cos \varphi(s, \beta) + y \sin \varphi(s, \beta) - r(s)) \, p(s, \beta) \, \mathrm{d}s \, \mathrm{d}\beta, \end{aligned} \tag{A.6}$$

where $r(s)$ and $\varphi(s, \beta)$ were defined in (A.3). We will next extend a common derivation of backproject then filter (BPF) tomographic reconstruction to accommodate a user defined weighting, and then change the coordinates from parallel-beam to fan-beam space.

---

[1] Practically speaking, the integral should be restricted to the field of view: $\sqrt{x^2 + y^2} \leq r_{\max}$, but this restriction would complicate analysis by introducing a shift variance into the problem, so we ignore it.

A.2 Parallel-Beam Grammian Analysis

When we analyze the local impulse response, we typically consider recentered local impulse functions. In this derivation we will start with an uncentered local impulse response, and center it at the end by removing a phase term in the frequency domain. The un-centered local impulse response of the Grammian operator is ,

$$\boldsymbol{\mathcal{P}}^*\boldsymbol{\mathcal{W}}\boldsymbol{\mathcal{P}}\boldsymbol{\delta}^j(x,y) = \int_0^\pi \int_{-\infty}^\infty \delta(x^j\cos(\phi) + y^j\sin(\phi) - r)$$

$$w(r,\phi)\delta(x\cos(\phi) + y\sin(\phi) - r)drd\phi.$$

Using the sampling property with the first δ, define

$$w^j(\phi) \triangleq w(r,\phi) = w(x^j\cos(\phi) + y^j\sin(\phi), \phi). \tag{A.7}$$

We denote the Fourier transform of $\delta^j(x,y)$ as $F^j(\rho,\Phi)$. Then, using the Fourier slice theorem,

$$\begin{aligned}\boldsymbol{\mathcal{P}}^*\boldsymbol{\mathcal{W}}\boldsymbol{\mathcal{P}}\boldsymbol{\delta}^j(x,y) &= \int_0^\pi \int_{-\infty}^\infty \int_{-\infty}^\infty F^j(\rho,\phi)e^{i2\pi\rho r}d\rho \\ &\quad w^j(\phi)\delta(x\cos\phi + y\sin\phi - r)drd\phi \\ &= \int_0^\pi \int_{-\infty}^\infty F^j(\rho,\Phi) \\ &\quad w^j(\phi)e^{i2\pi\rho(x\cos\phi + y\sin\phi)}d\rho d\phi.\end{aligned} \tag{A.8}$$

This is nearly the inverse Fourier transform in signed polar coordinates except for a ρ scale factor. Dividing by ρ,

$$\boldsymbol{\mathcal{P}}^*\boldsymbol{\mathcal{W}}\boldsymbol{\mathcal{P}}\boldsymbol{\delta}^j(x,y) = \int_0^\pi \int_{-\infty}^\infty \frac{w^j(\phi)F(\delta^j(x,y))}{\rho}$$

$$e^{i2\pi\rho(x\cos\phi + y\sin\phi)}d\rho d\phi.$$

The local impulse response is recentered, $h^j(x,y) = \boldsymbol{\mathcal{P}}^*\boldsymbol{\mathcal{W}}\boldsymbol{\mathcal{P}}\boldsymbol{\delta}^j(x+x^j, y+y^j)$ which eliminates the complex phase term $F^j(\rho,\phi)$. Then, the local frequency response is

$$H^j(\rho,\Phi) = \frac{w^j(\Phi)}{\rho}. \tag{A.9}$$

A.3 Fan Beam Grammian Analysis

The natural indexing for fan-beam data is arc-length s and angle β. The analogs to r,ϕ for parallel-beam systems are $r(s), \varphi(s,\beta)$ as defined in (A.3). The weighting expression $w(r,\phi)$ is indexed as $w(s,\beta)$ in fan-beam coordinates. We start by looking at the fan-beam projection in terms of the analogs to parallel-beam coordinates,

$$\begin{aligned}(\boldsymbol{\mathcal{P}}^*\boldsymbol{\mathcal{W}}\boldsymbol{\mathcal{P}}\boldsymbol{\delta}^j)(x,y) = \int_{-\beta_{\max}}^{\beta_{\max}}\int_{-s_{\max}}^{s_{\max}} & w(s,\beta) \\ & \delta(x^j\cos(\varphi(s,\beta)) + y^j\sin(\varphi(s,\beta)) - r(s)) \\ & \delta(x\cos(\varphi(s,\beta)) + y\sin(\varphi(s,\beta)) - r(s)) ds d\beta.\end{aligned}$$

We use the change of variables

$$r' = r(s) = D_{\mathrm{s0}}\sin\gamma(s) \tag{A.10}$$

$$\varphi' = \varphi(s,\beta) = \beta + \gamma(s) \tag{A.11}$$

as defined in (A.3) which has the corresponding Jacobian determinant

$$J(s) = |D_{\mathrm{s0}}\cos\gamma(s)||\dot{\gamma}(s)|. \tag{A.12}$$

Then,

$$\begin{aligned}(\boldsymbol{\mathcal{P}}^*\boldsymbol{\mathcal{W}}\boldsymbol{\mathcal{P}}\boldsymbol{\delta}^j)(x,y) = \int_0^{2\pi}\int_{-r_{\max}}^{r_{\max}} & w(r',\varphi') \\ & \delta(x^j\cos\varphi' + y^j\sin\varphi' - r') \\ & \delta(x\cos\varphi' + y\sin\varphi' - r')\frac{1}{J(s(r'))} dr d\varphi.\end{aligned}$$

In this expression,

$$\begin{aligned} w(r,\varphi) &\triangleq w(s(r),\beta(r,\varphi)) \\ s(r) &= \gamma^{-1}(\arcsin(r/D_{\mathrm{s0}})) \\ \beta(r,\varphi) &= \varphi - \arcsin(r/D_{\mathrm{s0}})\,. \end{aligned}$$

Using the sampling property of the first δ as in the parallel beam case,

$$w^j(\varphi') \triangleq w(x^j\cos\varphi' + y^j\sin\varphi', \varphi') \tag{A.13}$$

$$s^j(\varphi') \triangleq s(r')\Big|_{r'=x^j\cos\varphi'+y^j\sin\varphi'}\,. \tag{A.14}$$

Again, let $F^j(\rho,\Phi)$ denote the Fourier transform of $\delta^j(x,y)$. Then, using the Fourier slice theorem,

$$\begin{aligned} &(\boldsymbol{\mathcal{P}}^*\boldsymbol{\mathcal{W}}\boldsymbol{\mathcal{P}}\boldsymbol{\delta}^j)(x,y) \\ &= \int_{-r_{\max}}^{r_{\max}}\int_0^{2\pi}\int_{-\infty}^{\infty}\frac{F^j(\rho,\Phi)w^j(\varphi')}{J(s^j(\varphi'))}e^{i2\pi\rho r'} \\ &\qquad\qquad \delta(x\cos\varphi' + y\sin\varphi' - r')d\rho d\varphi' dr' \\ &= \int_0^{2\pi}\int_{-\infty}^{\infty}\frac{F^j(\rho,\Phi)w^j(\varphi)}{J(s^j(\varphi'))}e^{i2\pi\rho(x\cos\varphi'+y\sin\varphi')}d\rho d\varphi' \\ &= \int_0^{\pi}\int_{-\infty}^{\infty}\frac{F^j(\rho,\Phi)w^j(\varphi')}{J(s^j(\varphi'))}e^{i2\pi\rho(x\cos\varphi'+y\sin\varphi')}d\rho d\varphi' \\ &+ \int_0^{\pi}\int_{-\infty}^{\infty}\frac{F^j(\rho,\Phi)w^j(\varphi'')}{J(s^j(\varphi''))}e^{i2\pi\rho(x\cos\varphi''+y\sin\varphi'')}d\rho d\varphi'', \end{aligned}$$

where $\varphi'' = \varphi' + \pi$. This is similar to the parallel-beam derivation except that we have two integrals. We can convert each integral into the inverse Fourier transform as we did in the parallel-beam case and strip out the phase term $F^j(\rho,\Phi)$ by recentering the local impulse response. The local frequency responses of the Grammian operator is

$$H^j(\rho,\Phi) = \frac{\tilde{w}^j(\Phi)}{|\rho|}, \tag{A.15}$$

where,

$$\tilde{w}^j(\Phi) = \frac{1}{J(s^j(\Phi))}\left[w^j(\varphi)\Big|_{\varphi=\Phi} + w^j(\varphi)\Big|_{\varphi=\Phi+\pi} \right]. \tag{A.16}$$

Because of the absolute value function in (A.12), $J(s)$ is invariant to the π phase shift. For the case where we have uniform weighting, $w(s,\beta) = 1$ and therefore $\boldsymbol{W} = \boldsymbol{I}$, and (A.15) simplifies to

$$H^j(\rho,\Phi) = \frac{2}{J(s^j(\Phi))|\rho|}. \tag{A.17}$$

We use this equation in the calculation of a target local impulse response (3.19).

APPENDIX B

KKT proofs

In this appendix, we show that $\sqrt{d_2^2 + d_3^2} \leq d_1$, as claimed below (4.10) in §4.1.4. Squaring the integrals in (4.10), we have:

$$\begin{aligned} d_1^2 &= \frac{1}{\pi} \int_0^\pi \int_0^\pi w^j(X) w^j(Y) dX dY \\ d_2^2 &= \frac{1}{\pi} \int_0^\pi \int_0^\pi \cos(2X)\, w^j(X) \cos(2Y)\, w^j(Y) dX dY \\ d_3^2 &= \frac{1}{\pi} \int_0^\pi \int_0^\pi \sin(2X)\, w^j(X) \sin(2Y)\, w^j(Y) dX dY. \end{aligned}$$

In particular,

$$\begin{aligned} d_2^2 + d_3^2 &= \frac{1}{\pi} \int_0^\pi \int_0^\pi w^j(X) w^j(Y) \\ &\qquad [\cos(2X)\cos(2Y) + \sin(2X)\sin(2Y)] dX dY \\ &= \frac{1}{\pi} \int_0^\pi \int_0^\pi w^j(X) w^j(Y) \cos(2X - 2Y)\, dX dY. \end{aligned}$$

Thus, $d_2^2 + d_3^2 \leq d_1^2$ since $w^j(\Phi) \geq 0$ for all Φ and $\cos(2X - 2Y) \leq 1$ for all X and Y.

APPENDIX C

Streaks in Reconstructed Images

We attempted to gain more insight into why there were so many streaks in images reconstructed with edge preserving regularization so we decided to look at the quadratic case with the same phantom for some insight. We were disappointed to find that there were even more streaks surrounding bone in these images. An image reconstructed using certainty based regularization is displayed in Fig. C.1. These streaks are different than the ones we saw in the edge preserving case, Fig. 6.14-6.15. At this point we do not completely understand why these streaks are worse in the quadratic case than the edge preserving case. The streaks are present in regularized reconstruction using *AIMA* regularization, certainty based regularization, and even conventional regularization so we do not believe this to be the result of regularization design. The streaks disappear during unweighted reconstructions, Fig. C.2. We believe that our reconstruction algorithms have converged because further iterations do not seem to improve the cost function. Determining the cause of these streaks require further study, as well as why they are less apparent with edge-preserving regularization.

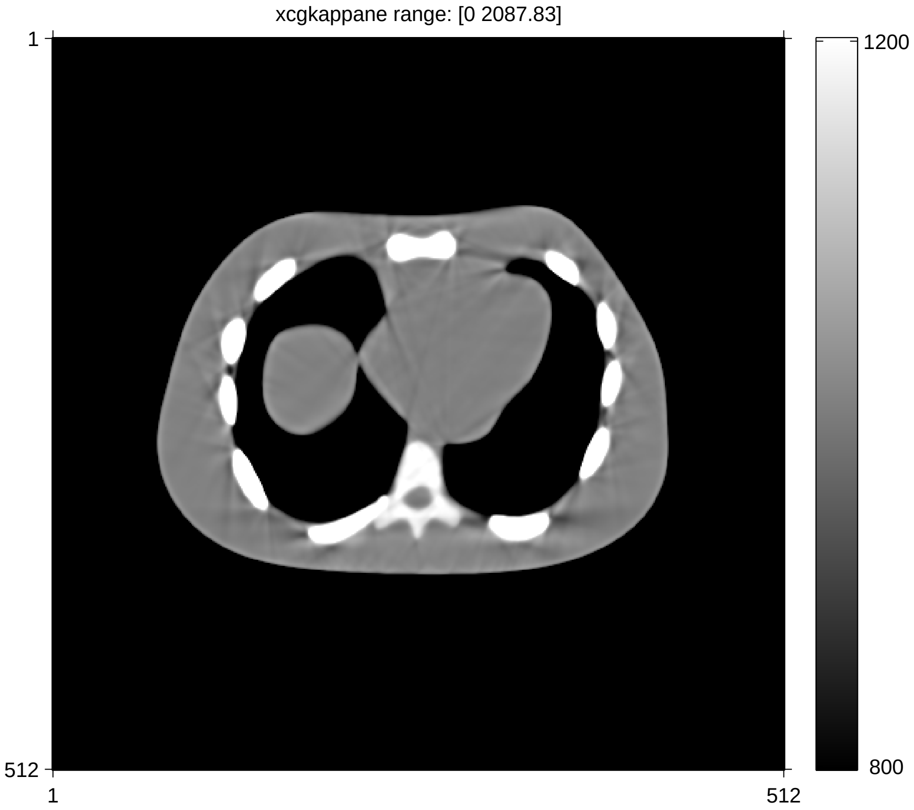


Figure C.1: Reconstruction of NCAT phantom using certainty based regularization.

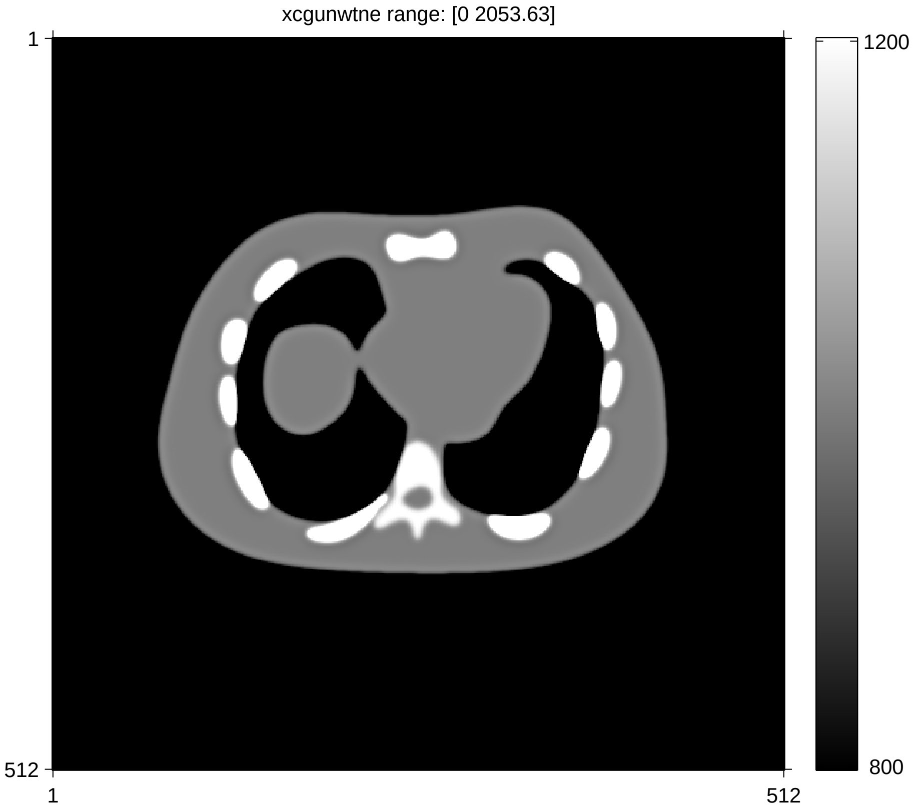


Figure C.2: Unweighted reconstruction of NCAT phantom using conventional regularization.

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] S. Ahn, J. A. Fessler, D. Blatt, and A. O. Hero. Convergent incremental optimization transfer algorithms: Application to tomography. *IEEE Trans. Med. Imag.*, 25(3):283–96, March 2006.

[2] S. Ahn and R. M. Leahy. Spatial resolution properties of nonquadratically regularized image reconstruction for PET. In *Proc. IEEE Intl. Symp. Biomed. Imag.*, pages 287–90, 2006.

[3] S. Ahn and R. M. Leahy. Analysis of resolution and noise properties of nonquadratically regularized image reconstruction methods for PET. *IEEE Trans. Med. Imag.*, 27(3):413–24, March 2008.

[4] E. Asma and R. M. Leahy. Mean and covariance properties of dynamic PET reconstructions from list-mode data. *IEEE Trans. Med. Imag.*, 25(1):42–54, January 2006.

[5] C. Bouman and K. Sauer. A generalized Gaussian image model for edge-preserving MAP estimation. *IEEE Trans. Im. Proc.*, 2(3):296–310, July 1993.

[6] L. T. Chang. A method for attenuation correction in radionuclide computed tomography. *IEEE Trans. Nuc. Sci.*, 25(1):638–643, February 1978.

[7] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud. Deterministic edge-preserving regularization in computed imaging. *IEEE Trans. Im. Proc.*, 6(2):298–311, February 1997.

[8] A. Chatziioannou, J. Qi, A. Annala, A. Moore, R. M. Leahy, and S. R. Cherry. Comparison of 3D MAP and FBP algorithms for image reconstruction in microPET. *IEEE Trans. Med. Imag.*, 19(5):507–12, May 2000.

[9] S. R. Cherry, M. Dahlbom, and E. J. Hoffman. 3D PET using a conventional multislice tomograph without septa. *J. Comp. Assisted Tomo.*, 15(4):655–68, July 1991.

[10] N. H. Clinthorne, T. S. Pan, P. C. Chiao, W. L. Rogers, and J. A. Stamos. Preconditioning methods for improved convergence rates in iterative reconstructions. *IEEE Trans. Med. Imag.*, 12(1):78–83, March 1993.

[11] B. De Man and S. Basu. Distance-driven projection and backprojection. In *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, volume 3, pages 1477–80, 2002.

[12] I. A. Elbakri and J. A. Fessler. Efficient and accurate likelihood for iterative image reconstruction in X-ray computed tomography. In *Proc. SPIE 5032, Medical Imaging 2003: Image Proc.*, pages 1839–50, 2003.

[13] J. A. Fessler. Penalized weighted least-squares image reconstruction for positron emission tomography. *IEEE Trans. Med. Imag.*, 13(2):290–300, June 1994.

[14] J. A. Fessler. Hybrid Poisson/polynomial objective functions for tomographic image reconstruction from transmission scans. *IEEE Trans. Im. Proc.*, 4(10):1439–50, October 1995.

[15] J. A. Fessler. Resolution properties of regularized image reconstruction methods. Technical Report 297, Comm. and Sign. Proc. Lab., Dept. of EECS, Univ. of Michigan, Ann Arbor, MI, 48109-2122, August 1995.

[16] J. A. Fessler. Mean and variance of implicitly defined biased estimators (such as penalized maximum likelihood): Applications to tomography. *IEEE Trans. Im. Proc.*, 5(3):493–506, March 1996.

[17] J. A. Fessler. Approximate variance images for penalized-likelihood image reconstruction. In *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, volume 2, pages 949–52, 1997.

[18] J. A. Fessler. Analytical approach to regularization design for isotropic spatial resolution. In *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, volume 3, pages 2022–6, 2003.

[19] J. A. Fessler. Matlab tomography toolbox, 2004. Available from `http://www.eecs.umich.edu/∼fessler`.

[20] J. A. Fessler and S. D. Booth. Conjugate-gradient preconditioning methods for shift-variant PET image reconstruction. *IEEE Trans. Im. Proc.*, 8(5):688–99, May 1999.

[21] J. A. Fessler, N. H. Clinthorne, and W. L. Rogers. Regularized emission image reconstruction using imperfect side information. *IEEE Trans. Nuc. Sci.*, 39(5):1464–71, October 1992.

[22] J. A. Fessler and W. L. Rogers. Uniform quadratic penalties cause nonuniform image resolution (and sometimes vice versa). In *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, volume 4, pages 1915–9, 1994.

[23] J. A. Fessler and W. L. Rogers. Spatial resolution properties of penalized-likelihood image reconstruction methods: Space-invariant tomographs. *IEEE Trans. Im. Proc.*, 5(9):1346–58, September 1996.

[24] P. J. Green. Bayesian reconstructions from emission tomography data using a modified EM algorithm. *IEEE Trans. Med. Imag.*, 9(1):84–93, March 1990.

[25] T. Hebert and R. Leahy. A generalized EM algorithm for 3-D Bayesian reconstruction from Poisson data using Gibbs priors. *IEEE Trans. Med. Imag.*, 8(2):194–202, June 1989.

[26] E. J. Hoffman, S. C. Huang, D. Plummer, and M. E. Phelps. Quantitation in positron emission tomography: 6 effect of nonuniform resolution. *J. Comp. Assisted Tomo.*, 6(5):987–99, October 1982.

[27] I-T. Hsiao, A. Rangarajan, and G. Gindi. A new convex edge-preserving median prior with applications to tomography. *IEEE Trans. Med. Imag.*, 22(5):580–5, May 2003.

[28] P. M. Joseph. The influence of modulation transfer function shape on CT image quality. *Radiology*, 145(1):179–85, October 1982.

[29] B. Karuta and R. Lecomte. Effect of detector weighting functions on the point spread function of high-resolution PET tomographs. *IEEE Trans. Med. Imag.*, 11(3):379–85, September 1992.

[30] L. A. Kunyansky. A new SPECT reconstruction algorithm based on the Novikov explicit inversion formula. *Inverse Prob.*, 17(2):293–306, April 2001.

[31] K. Lange. Convergence of EM image reconstruction algorithms with Gibbs smoothing. *IEEE Trans. Med. Imag.*, 9(4):439–46, December 1990. Corrections, T-MI, 10:2(288), June 1991.

[32] C. L. Lawson and R. J. Hanson. *Solving least squares problems*. Prentice-Hall, 1974.

[33] R. Leahy and X. H. Yan. Statistical models and methods for PET image reconstruction. In *Proc. of Stat. Comp. Sect. of Amer. Stat. Assoc.*, pages 1–10, 1991.

[34] S-J. Lee, A. Rangarajan, and G. Gindi. Bayesian image reconstruction in SPECT using higher order mechanical models as priors. *IEEE Trans. Med. Imag.*, 14(4):669–80, December 1995.

[35] Q. Li, E. Asma, J. Qi, J. R. Bading, and R. M. Leahy. Accurate estimation of the Fisher information matrix for the PET image reconstruction problem. *IEEE Trans. Med. Imag.*, 23(9):1057–64, September 2004.

[36] J. S. Liow and S. C. Strother. The convergence of object dependent resolution in maximum likelihood based tomographic image reconstruction. *Phys. Med. Biol.*, 38(1):55–70, January 1993.

[37] C. E. Metz and X. Pan. A unified analysis of exact methods of inverting the 2-D exponential Radon transform with implications for noise control in SPECT. *IEEE Trans. Med. Imag.*, 14(4):643–58, December 1995.

[38] J. Nuyts, D. Beque, P. Dupont, and L. Mortelmans. A concave prior penalizing relative differences for maximum-a-posteriori reconstruction in emission tomography. *IEEE Trans. Nuc. Sci.*, 49(1-1):56–60, February 2002.

[39] J. Nuyts and J. A. Fessler. Comparison between penalized-likelihood image reconstruction for uniform spatial resolution and post-smoothed ML-EM. In *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, volume 2, pages 895–9, 2002.

[40] J. Nuyts and J. A. Fessler. A penalized-likelihood image reconstruction method for emission tomography, compared to post-smoothed maximum-likelihood with matched spatial resolution. *IEEE Trans. Med. Imag.*, 22(9):1042–52, September 2003.

[41] A. V. Oppenheim and R. W.. Schafer. *Discrete-time signal processing*. Prentice-Hall, New York, 1989.

[42] R. Pan and S. J. Reeves. Efficient Huber-Markov edge-preserving image restoration. *IEEE Trans. Im. Proc.*, 15(12):3728–35, December 2006.

[43] V. Y. Panin, G. L. Zeng, and G. T. Gullberg. Total variation regulated EM algorithm. *IEEE Trans. Nuc. Sci.*, 46(6):2202–10, December 1999.

[44] Image Processing, Dept. of Diagnostic Radiology Analysis Group, Electrical Engineering, and Yale University the Yale School of Medecine. The zubal phantom. Available Online: http://noodle.med.yale.edu/zubal/.

[45] J. Qi. Investigation of lesion detection in map reconstruction with non-gaussian priors. In *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, volume 3, pages 1704–1708, 2005.

[46] J. Qi. Comparison of lesion detection and quantification in map reconstruction with gaussion and non-gaussian priors. *Int. J. Biomed. Imag.*, 2006. Available Online: http://www.hindawi.com/GetArticle.aspx?doi=10.1155/IJBI/2006/87567.

[47] J. Qi. Comparison of statistical reconstructions with isotropic and anisotropic resolution in PET. *IEEE Trans. Nuc. Sci.*, 53(1):147–51, February 2006.

[48] J. Qi and R. H. Huesman. Theoretical study of lesion detectability of MAP reconstruction using computer observers. *IEEE Trans. Med. Imag.*, 20(8):815–22, August 2001.

[49] J. Qi and R. H. Huesman. Penalized maximum-likelihood image reconstruction for lesion detection. *Phys. Med. Biol.*, 51(16):4017–30, August 2006.

[50] J. Qi and R. M. Leahy. Resolution and noise properties of MAP reconstruction for fully 3D PET. In *Proc. Intl. Mtg. on Fully 3D Image Recon. in Rad. and Nuc. Med*, pages 35–9, 1999.

[51] J. Qi and R. M. Leahy. A theoretical study of the contrast recovery and variance of MAP reconstructions from PET data. *IEEE Trans. Med. Imag.*, 18(4):293–305, April 1999.

[52] J. Qi and R. M. Leahy. Resolution and noise properties of MAP reconstruction for fully 3D PET. *IEEE Trans. Med. Imag.*, 19(5):493–506, May 2000.

[53] J. Radon. On the determination of functions from their integrals along certain manifold. *Berichte Sächs. Akad. Wiss. (Leipzig)*, 69:262–78, 1917. Uber die Bestimmung von Funktionen durch ihre Intergralwerte Langs gewisser Manningfultigkeiten.

[54] A. J. Rockmore and A. Macovski. A maximum likelihood approach to emission image reconstruction from projections. *IEEE Trans. Nuc. Sci.*, 23:1428–32, 1976.

[55] A. J. Rockmore and A. Macovski. A maximum likelihood approach to transmission image reconstruction from projections. *IEEE Trans. Nuc. Sci.*, 24(3):1929–35, June 1977.

[56] D. Schmitt, B. Karuta, C. Carrier, and R. Lecomte. Fast point spread function computation from aperture functions in high-resolution positron emission tomography. *IEEE Trans. Med. Imag.*, 7(1):2–12, March 1988.

[57] W. Segars. *Development and Application of the New Dynamic NURBS-based Cardiac-Torso (NCAT) Phantom*. PhD thesis, Dept. Biomedical Engineering, University of North Carolina, 2001.

[58] H. Shi and J. A. Fessler. Quadratic regularization design for 3d cylindrical PET. In *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, volume 4, pages 2301–5, 2005.

[59] H. Shi and J. A. Fessler. Quadratic regularization design for fan beam transmission tomography. In *Proc. SPIE 5747, Medical Imaging 2005: Image Proc.*, pages 2023–33, 2005.

[60] H. R. Shi and J. A. Fessler. Quadratic regularization design for 2D CT. *IEEE Trans. Med. Imag.*, 2007. Submitted as TMI-2007-0296.

[61] D. L. Snyder, M. I. Miller, L. J. Thomas, and D. G. Politte. Noise and edge artifacts in maximum-likelihood reconstructions for emission tomography. *IEEE Trans. Med. Imag.*, 6(3):228–38, September 1987.

[62] E. J. Soares, C. L. Byrne, S. J. Glick, C. R. Appledorn, and M. A. King. Implementation and evaluation of and analytical solution to the photon attenuation and nonstationary resolution reconstruction problem in SPECT. *IEEE Trans. Nuc. Sci.*, 40(4):1231–7, August 1993.

[63] S. Srivastava. *Accelerated statistical image reconstruction algorithms and simplified cost functions for X-ray computed tomography*. PhD thesis, Univ. of Michigan, Ann Arbor, MI, 48109-2122, Ann Arbor, MI, 2008.

[64] S. Srivastava and J. A. Fessler. Penalized likelihood transmission image reconstruction: Unconstrained monotonic algorithms. In *Proc. IEEE Intl. Symp. Biomed. Imag.*, pages 748–51, 2004.

[65] J. W. Stayman and J. A. Fessler. Regularization for uniform spatial resolution properties in penalized-likelihood image reconstruction. *IEEE Trans. Med. Imag.*, 19(6):601–15, June 2000.

[66] J. W. Stayman and J. A. Fessler. Nonnegative definite quadratic penalty design for penalized-likelihood reconstruction. In *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, volume 2, pages 1060–3, 2001.

[67] J. W. Stayman and J. A. Fessler. Compensation for nonuniform resolution using penalized-likelihood reconstruction in space-variant imaging systems. *IEEE Trans. Med. Imag.*, 23(3):269–84, March 2004.

[68] J. W. Stayman and J. A. Fessler. Efficient calculation of resolution and covariance for fully-3D SPECT. *IEEE Trans. Med. Imag.*, 23(12):1543–56, December 2004.

[69] J-B. Thibault, C. A. Bouman, K. D. Sauer, and J. Hsieh. A recursive filter for noise reduction in statistical iterative tomographic imaging. In *Proc. SPIE 6065, Computational Imaging IV*, page 60650X, 2006.

[70] J-B. Thibault, K. Sauer, C. Bouman, and J. Hsieh. A three-dimensional statistical approach to improved image quality for multi-slice helical CT. *Med. Phys.*, 34(11):4526–44, November 2007.

[71] A. Tikhonov and V. Arsenin. *Solution of ill-posed problems*. Wiley, New York, 1977.

[72] E. Veklerov and J. Llacer. Stopping rule for the MLE algorithm based on statistical hypothesis testing. *IEEE Trans. Med. Imag.*, 6(4):313–9, December 1987.

[73] P. R. G. Virador, W. W. Moses, R. H. Huesman, and J. Qi. 3-D reconstruction in PET cameras with irregular sampling and depth of interaction. *IEEE Trans. Nuc. Sci.*, 48(4-2):1524–9, August 2001.

[74] D. W. Wilson and B. M. W. Tsui. Spatial resolution properties of FB and ML-EM reconstruction methods. In *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, volume 2, pages 1189–93, 1993.

[75] W. Xia, R. M. Lewitt, and P. R. Edholm. Fourier correction for spatially variant collimator blurring in SPECT. *IEEE Trans. Med. Imag.*, 14(1):100–15, March 1995.

[76] R. Yao, J. Seidel, C. A. Johnson, M. E. Daube-Witherspoon, M. V. Green, and R. E. Carson. Performance characteristics of the 3-D OSEM algorithm in the reconstruction of small animal PET images. *IEEE Trans. Med. Imag.*, 19(8):798–804, August 2000.

[77] Y. Zhang-O'Connor and J. A. Fessler. Fast predictions of variance images for fan-beam transmission tomography with quadratic regularization. *IEEE Trans. Med. Imag.*, 26(3):335–46, March 2007.

[78] G. Zubal, G. Gindi, M. Lee, C. Harrell, and E. Smith. High resolution anthropomorphic phantom for Monte Carlo analysis of internal radiation sources. In *IEEE Symposium on Computer-Based Medical Systems*, pages 540–7, 1990.

ABSTRACT

Fast Regularization Design for Tomographic Image Reconstruction for Uniform and
Isotropic Spatial Resolution

by

Hugo R. Shi

Chair: Jeffrey A. Fessler

Statistical methods for tomographic image reconstruction have improved noise and spatial resolution properties that may improve image quality in X-ray CT and PET. Final converged solutions from maximum likelihood (ML) and weighted least squares (WLS) reconstruction are often extremely noisy due to the ill conditioned nature of the system. One can stop the iterative algorithm before convergence and before images become too noisy, however this results in non-uniform and anisotropic spatial resolution because resolution uniformity and isotropy improve with successive iterations. Alternatively, one can run the iterative algorithm to completion and post-filter the resulting noise, however, this often requires a large number of iterations. Instead we use penalized likelihood (PL) and penalized weighted least squares (PWLS) with a roughness penalty to regularize the problem which filters out noise, and leads to faster convergence. Unfortunately, interactions between the weightings, which are implicit in PL methods and explicit in PWLS methods, and conventional quadratic regularization lead to nonuniform and anisotropic spatial resolution. Previous work focuses on matrix algebra methods to design data-dependent, shift variant regularizers that improve resolution uniformity. This thesis develops fast analytical regularization design methods for 2D fan-beam X-ray CT imaging, for which parallel-beam tomography is a special case. This thesis uses continuous space analogs to greatly simplify the regularization design problem which yields a mostly analytical solution for efficient computation. This thesis extends regularization design to 3D systems using a computationally efficient iterative approach. Finally, this thesis explores using 2D regularization with z-dimension post-reconstruction denoising. This is an attempt to combine the improved XY isotropy associated with 2D regularization design, and the computational efficiency of the mostly analytical solution and use it for 3D geometries. The spatial resolution and noise properties of this method is analyzed for quadratic regularizers. Simulation results have also been performed using non-quadratic edge-preserving regularizers which show that, though this method has potential, more work needs to be done to ensure that the spatial resolution and noise properties of this method are desirable.