

Statistical Image Reconstruction and Motion Estimation for Image-Guided Radiotherapy

by

Yong Long

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Electrical Engineering: Systems)
in The University of Michigan
2011

Doctoral Committee:

Professor Jeffrey A. Fessler, Co-Chair
Professor James M. Balter, Co-Chair
Professor Alfred O. Hero III
Professor Charles R. Meyer
Research Professor Neal H. Clinthorne

© Yong Long 2011
All Rights Reserved

To my parents, husband and daughter.

ACKNOWLEDGEMENTS

This thesis would have been impossible without my advisors, Prof. Jeff Fessler and Prof. James Balter. Working with them has been enjoyable and rewarding. They have understood difficulties from both my research and personal life, have always been supportive and willing to provide extra help when needed. I especially thank Jeff for his guidance, support and patience before I was formally enrolled at the University of Michigan. His deep insight and passion for research and dedication to his students are amazing. James always noticed my progress and encouraged me to accomplish more. I especially thank him for giving me freedom to pursue my research interest and easy access to clinical data, and setting our meetings at EECS for my convenience when I was pregnant.

I would like to express my gratitude to Prof. Alfred Hero, Prof. Charles Meyer and Prof. Neal Clinthorne for their expertise and valuable comments on this thesis. It was such an honor to have them as my thesis committee members.

I would like to thank my colleagues and friends for making my stay in Ann Arbor enjoyable. Special thanks goes to Dan Ruan who has been a “big” sister to me over the years. From helping me with course problems, to discussing research topics, to being my favorite shopping companion, her friendship and support is incredible. I thank Rongping Zeng, Yingying Zhang-O’Connor, Somesh Srivastava and Se Young Chun for their guidance when I was a beginner.

I am deeply indebted to my parents, Youbin Long and Shihui Yang, for their unconditional

love, support, encouragement and believing in me throughout my life. I cannot ask for more from them, as they are the best parents. They have traveled from far away to take care of me during my late pregnancy and postpartum recovery, and to help attending my baby daughter. Without their help, I never would have finished this thesis when my daughter is only about four-month old. My dear husband, Weiming, has been a great source of love, support and comfort over all the years. I am very grateful to his staying in Ann Arbor to be with me after his graduation in 2009. I deeply appreciate my daughter, Chengwen, for the unqualified love, joy, pride and excitement she has brought to me.

This work was financially supported in part by NIH grant P01-CA59827 and Barbour fellowship at the University of Michigan.

TABLE OF CONTENTS

DEDICATION	ii
ACKNOWLEDGEMENTS	iii
LIST OF FIGURES	viii
LIST OF TABLES	xi
LIST OF APPENDICES	xii
ABSTRACT	xiii
CHAPTER	
1. Introduction	1
1.1 Image-Guided Radiotherapy (IGRT)	1
1.2 Image Reconstruction for X-Ray Computed Tomography (CT)	2
1.3 Motion Estimation for IGRT	3
1.4 Contributions and Outline	4
2. Background	7
2.1 Principles of X-Ray CT	7
2.1.1 General Measurement Model	7
2.1.2 Cone-Beam Geometry	9
2.1.3 Image Reconstruction Methods	12
2.1.4 Beam-Hardening Correction	19
2.2 2D-3D Image Registration	25
2.2.1 Similarity Measures	27
2.2.2 Nonrigid Deformation Model	27
2.2.3 Regularization of Image Registration	30
2.3 Optimization Transfer Principle	30

3. Three-Dimensional Forward and Back-Projection Methods	32
3.1 Cone-Beam Projection	36
3.1.1 Cone-Beam System Model	36
3.1.2 Footprints of Voxel Basis Functions	38
3.2 Separable Footprint (SF) Projector	39
3.2.1 Amplitude Approximation Methods	42
3.2.2 SF Projector with Trapezoid/Rectangle Function (SF-TR)	44
3.2.3 SF Projector with Trapezoid/Trapezoid Function (SF-TT)	46
3.2.4 Implementation of SF projector	47
3.2.5 SF Compared with Distance-Driven (DD) Projector	50
3.3 Simulation Results	51
3.3.1 Forward and Back-Projector as Single Modules	52
3.3.2 Forward and Back-Projector within Iterative Reconstruction	56
3.4 Conclusion	62
4. Two-Material Decomposition from A Single CT Scan Using Statistical Image Reconstruction	64
4.1 Differential Filtration	68
4.1.1 Split Filtration	69
4.1.2 Bow-tie Filtration	71
4.2 Penalized Weighted Least-Squares (PWLS) Reconstruction	71
4.3 Optimization Algorithm	73
4.3.1 Separable Quadratic Surrogate of Sinogram Vector	74
4.3.2 Separable Quadratic Surrogate of Image Vector	75
4.4 Simulation Results	76
4.5 Discussion and Conclusion	80
5. Accuracy Estimation for Projection-to-Volume Targeting During Rotational Therapy: A Feasibility Study	91
5.1 Materials and Methods	92
5.1.1 Projection-to-Volume Registration	92
5.1.2 Investigating the Influence of Rotational Arc Length	94
5.2 Results and Discussion	97
5.2.1 Noiseless Projection Views	98
5.2.2 Noisy Projection Views	105
5.3 Conclusion	106
6. Conclusion and Future Work	108

6.1	Summary	108
6.2	Future Work	110
APPENDICES		114
BIBLIOGRAPHY		135

LIST OF FIGURES

<u>Figure</u>		
2.1	Schematic diagram of a X-ray scanner.	8
2.2	Axial cone-beam flat-detector geometry.	9
2.3	Energy-dependent mass attenuation coefficients $\beta_i(\mathcal{E})$ for bone, fat and water. . .	20
3.1	Exact footprint functions $q(s, t; \beta; \vec{n})$ and their profiles for 1mm^3 voxels centered at the origin (<i>left</i>), (100, 150, 15) mm (<i>center</i>) and (93, 93, 93) mm (<i>right</i>). . . .	40
3.2	Maximum error comparison between the forward DD, SF-TR and SF-TT projector for a voxel centered at the origin (<i>left</i>) and a voxel centered at (100, 150, -100) mm (<i>right</i>).	54
3.3	RMS error comparison between the forward DD, SF-TR and SF-TT projector for a voxel centered at the origin (<i>left</i>) and a voxel centered at (100, 150, -100) mm (<i>right</i>).	54
3.4	Shepp-Logan digital phantoms in Hounsfield units. The first, second and third columns show axial, coronal and sagittal views respectively.	58
3.5	Axial views of FOV images $\hat{\mathbf{x}}_{\text{FOV}}^{\text{SF-TR}}$ and $\hat{\mathbf{x}}_{\text{FOV}}^{\text{DD}}$ reconstructed by the iterative method (PWLS-CG) using the SF and DD method respectively. <i>Left</i> : SF-TR projector; <i>Right</i> : DD projector.	58
3.6	Axial views of ROI images $\hat{\mathbf{x}}_{\text{ROI}}^{\text{SF-TR}}$ and $\hat{\mathbf{x}}_{\text{ROI}}^{\text{DD}}$ reconstructed by the iterative method (PWLS-CG) using the SF-TR and DD method respectively. <i>Left</i> : SF-TR projector; <i>Right</i> : DD projector.	59
4.1	Schematic diagram of a fan-beam CT scanner with two positions of X-ray source and split-filter assembly.	69

4.2	<i>Left</i> : Spectra produced by a split filter for a simulated spectrum with X-ray tube voltage of 80 kVp. The split filter consists of 2 mm aluminum producing "soft" spectra and 200 μm molybdenum producing "hard" spectra. The effective energies are 49 keV and 58 keV respectively. <i>Right</i> : Sample spectra at four fan angles screened by a body bow-tie filter. γ denotes the fan angle. The effective energies are 49, 51, 53 and 56 keV for the spectra at 26° , 14° , 10° and 0° respectively. . . .	70
4.3	True soft tissue, bone and density map (soft tissue + bone) of a NCAT chest phantom.	79
4.4	Logarithm sinogram estimates \hat{f}_{im} using the simulated split filter (<i>Left</i>) and bow-tie filter (<i>Right</i>).	80
4.5	Water and bone corrected FBP and their absolute error images using split and bow-tie filtration.	81
4.6	Soft tissue images and their absolute errors reconstructed by the JS-FBP and proposed PWLS method using split and bow-tie filtration.	82
4.7	Bone images and their absolute errors reconstructed by the JS-FBP and proposed PWLS method using split and bow-tie filtration.	83
4.8	Density maps (the sum of soft tissue and bone images) reconstructed by the JS-FBP and PWLS method using split and bow-tie filtration. The images are shown in a more sensitive viewing window ($[0.8, 1.2] \text{ g/cm}^3$).	84
4.9	Linear attenuation coefficient images at 511 keV reconstructed by the JS-FBP and PWLS method using split and bow-tie filtration. The range of true linear attenuation coefficients was $[0.08, 0.12] \text{ cm}^{-1}$. The images are shown in a more sensitive viewing window ($[0.08, 0.12] \text{ cm}^{-1}$) to highlight the beam-hardening effects.	85
4.10	Vertical profiles through the true and estimated soft tissue, bone components and the linear attenuation coefficient images by the JS-FBP and PWLS method using split (<i>left</i>) and bow-tie (<i>right</i>) filtration. The green line in the true density image in Fig. 4.11 shows profile locations.	86
4.11	Vertical profiles of PWLS method using the simulated bow-tie and split filter. The green line in the true linear attenuation image at 511 keV shows the location of profiles in this figure and Fig. 4.10.	87

5.1	Axial (<i>left</i>), sagittal (<i>center</i>) and coronal views (<i>right</i>) of reference, target volumes and reference volume within a ROI overlaid with example radial annuli for analysis.	95
5.2	Schematic diagram of limited-angle scans centered at 0° , 45° and 90° . p and o are the axes along and orthogonal to the ray connecting the X-ray source and the detector center for the limited-angle scan centered at 45°	97
5.3	Absolute errors (mm) of tumor center shift.	99
5.4	Mean absolute errors (mm) of estimated deformation within ROI. <i>Left</i> : Angular center is $\beta = 0^\circ$. <i>Center</i> : Angular center is $\beta = 45^\circ$. <i>Right</i> : Angular center is $\beta = 90^\circ$	102
5.5	Maximum absolute errors (mm) of estimated deformation within ROI. <i>Left</i> : Angular center is $\beta = 0^\circ$. <i>Center</i> : Angular center is $\beta = 45^\circ$. <i>Right</i> : Angular center is $\beta = 90^\circ$	102
5.6	Standard deviation of absolute errors (mm) of estimated deformation within ROI. <i>Left</i> : Angular center is $\beta = 0^\circ$. <i>Center</i> : Angular center is $\beta = 45^\circ$. <i>Right</i> : Angular center is $\beta = 90^\circ$	103
5.7	Histograms in the o direction (<i>left</i>), p direction (<i>center</i>) and z direction (<i>right</i>) within the ROI for the small deformation case.	103
5.8	Mean (<i>left</i>) and maximum (<i>right</i>) absolute errors of estimated displacements within ROI for a limited-angle scan with range of 18° centered at 0° . The horizontal axis denotes radial distance to the tumor center. The errors were calculated among points within cylindrical shells centered at the tumor center.	104
B.1	$f(z)$ when $d_n = 0.5$	129
B.2	$f(\alpha)$ for $a_n = 1$, $b_n = 1.1$ and $c_n = 4.3$. The minimum is at $\alpha = 2.65$	132

LIST OF TABLES

Table

3.1	Pseudo-code for the SF-TR forward projector with the A1 amplitude method (SF-TR-A1) and the A2 method (SF-TR-A2).	48
3.2	Speed comparison of DD, SF-TR and SF-TT forward and back projectors.	55
4.1	RMS error comparison of reconstructed images by the JS-FBP and proposed PWLS method using split and bow-tie filtration. The units are 10^{-2}g/cm^{-3} for the soft tissue, bone and density images, and 10^{-3}cm^{-1} for the linear attenuation coefficient images at 511 keV.	79
5.1	Absolute errors (mm) for angular ranges of 360° and $180^\circ([-90^\circ, 90^\circ])$	99
5.2	Mean, maximum and standard deviation σ of absolute errors of estimated displacements with angular ranges of $180^\circ([-90^\circ, 90^\circ])$ and 360°	101
5.3	Absolute errors (mm) of tumor center shifts for a limited-angle scan with range of 24° centered at 45° of the small deformation case. I^a stands for the case of no noise, I^b stands for the case when $I_{m,n} = 10^5$ counts per ray, and I^c stands for the case when $I_{m,n} = 10^4$ counts per ray.	106
5.4	Mean, maximum and standard deviation σ of absolute errors of estimated displacements within ROI for a limited-angle scan with range of 24° centered at 45° of the small deformation case. I^a stands for the case of no noise, I^b stands for the case when $I_{m,n} = 10^5$ counts per ray, and I^c stands for the case when $I_{m,n} = 10^4$ counts per ray.	106

LIST OF APPENDICES

Appendices

A.	Surrogate Function Design	115
A.1	Local Properties of Beam-Hardening Function	115
A.2	Curvature for Quadratic Surrogate	117
A.3	Separable Quadratic Surrogates Based on Additive Convexity Trick	117
A.3.1	Curvature of Diagonal Matrix	120
B.	Optimal Curvature for Two-Dimensional Quadratic Surrogates	121
B.1	Definition of An Optimal Curvature Matrix	122
B.2	Majorization Constraint	123
B.2.1	Diagonal curvature matrix	124
B.2.2	Curvature of Constant Times the Identity Matrix	126
B.3	Minimum Spectral Radius Criterion	127
B.3.1	Diagonal Curvature Matrix	127
B.3.2	Curvature of Constant Times the Identity Matrix	129
B.4	Optimal Curvatures of Minimum Spectral Radius Criterion	132
B.4.1	Diagonal Curvature Matrix	132
B.4.2	Curvature of Constant Times the Identity Matrix	133

ABSTRACT

Statistical Image Reconstruction and Motion Estimation for Image-Guided Radiotherapy

by

Yong Long

Co-Chairs: Jeffrey A. Fessler and James M. Balter

Image reconstruction and motion estimation are very important for image-guided radiotherapy (IGRT). Three-dimensional reconstruction of patient anatomy using X-ray computed tomography (CT) allows identification of the location of a tumor prior to treatment. The locations of tumors may change during actual treatment due to movement such as respiratory motion. Motion estimation helps optimize the accuracy and precision of radiotherapy so that more of the normal surrounding tissue can be spared. This dissertation addresses several important issues related to these two core components of IGRT.

Firstly, we developed two new separable footprint (SF) projector methods for X-ray cone-beam CT. The SF projectors approximate the voxel footprint functions as 2D separable functions. The SF-TR projector uses trapezoid functions in the transaxial direction and rectangular functions in the axial direction, whereas the SF-TT projector uses trapezoid functions in both directions. Both SF projector methods are more accurate than the distance-driven (DD) projector, which is a

current state-of-the-art method in the field. The SF-TT projector is more accurate than the SF-TR projector for rays associated with large cone angles. In addition, the SF-TR projector has similar computation speed with the DD projector and the SF-TT projector is about two times slower.

Secondly, we proposed a statistical penalized weighted least-squares (PWLS) method with edge-preserving regularization to reconstruct two basis materials from a single-energy CT scan acquired with differential filtration, such as a split filter or a bow-tie filter. It requires only the use of suitable filters between the X-ray tube and the patient. For both filtration methods, the proposed PWLS method reconstructed soft tissue and bone images with lower RMS errors, reduced the beam-hardening artifacts much more effectively and produced lower noise, as compared with the traditional non-iterative Joseph and Spital method.

Thirdly, we conducted an objective characterization of the influence of rotational arc length on accuracy of motion estimation for projection-to-volume targeting during rotational therapy. Simulations illustrate the potential accuracy of limited-angle projection-to-volume alignment. Registration accuracy can be sensitive to angular center, tends to be lower along direction of the projection set, and tends to decrease away from the rotation center.

CHAPTER 1

Introduction

1.1 Image-Guided Radiotherapy (IGRT)

Radiotherapy is a treatment method for cancer and other disease. About 40% of cancer patients have radiotherapy as part of their treatment [118]. Radiotherapy uses “radiation”, usually X-rays, to destroy cancer tissue. The radiation dosage is specially tailored for each patient, according to the position, size and shape of the tumor, the patient’s body shape in that area and the patient’s general health [119]. Since its first use to treat cancer, radiotherapy has continued to evolve with numerous changes and developments to accomplish its ultimate goal of effective destruction of cancer tissue while delivering a minimal dose of radiation to adjacent healthy tissues [52].

X-ray Computed Tomography (CT) scanners were introduced into clinic practice in the early 1970s [7]. This machine takes a series of 2D projection radiographs of the patient’s tumor and internal organs at different angles while the X-ray source rotates around the patient. These radiographs are put together by reconstruction methods to obtain a 3D image of the anatomy of the body being scanned. This 3D image of spatial distribution of attenuation coefficients shows where the tumor is, how big it is and how close major body organs are to the treatment area, which enables physicians to plan radiotherapy beams more precisely [116].

However, the locations of tumors may change during treatment due to constant movement such

as respiratory motion [12, 138]. Some tumors in the thoracic and abdominal regions may move as much as 3 cm peak-to-peak during radiotherapy treatment. Techniques such as Intensity Modulated Radiotherapy (IMRT) [113] minimize the dose to surrounding normal critical structures, using beam shaping devices such as “multi-leaf” collimators (lead shutters) to allow shape and intensity changes of radiation beams during treatment. Image-guided radiotherapy (IGRT) [117] is evolving to provide improved planar or X-ray volume imaging during treatment. The movement of the target tumor and critical organs can be estimated by registration between a reference image taken before treatment and these image data obtained during treatment. The local configuration information enables doctors to target the radiation beam at tumors precisely, and consequently spare the normal surrounding tissue.

1.2 Image Reconstruction for X-Ray Computed Tomography (CT)

Iterative statistical methods for 3D tomographic image reconstruction [34, 74, 147] offer numerous advantages such as the potential for improved image quality and reduced dose, as compared to the conventional methods such as filtered back-projection (FBP) [30]. They are based on models for measurement statistics and physics, and can easily incorporate prior information, the system geometry and the detector response.

The main disadvantage of statistical reconstruction methods is the longer computation time of iterative algorithms that are usually required to minimize certain cost functions. For most iterative reconstruction methods, each iteration requires one forward projection and one back-projection, where the forward projection is roughly a discretized evaluation of the Radon transform, and the back-projector is the adjoint of the forward projector. These operations are the primary computational bottleneck in iterative reconstruction methods, particularly in 3D image

reconstruction. A variety of methods for 3D forward and back-projection have been proposed [9, 14, 26, 38, 79–81, 93, 120, 167]. All of them provide some compromise between computational complexity and accuracy. We discuss projector methods in Chapter 3.

Material attenuation coefficients depend on the energy of the incident photons. An X-ray beam in clinical practice is usually composed of individual photons with a wide range of energies, and each photon energy is attenuated differently by the object. This energy dependence causes artifacts in reconstructed images, such as beam-hardening artifacts [18]. The beam becomes “harder” (its mean energy increases) as it passes through the object because the lower-energy photons are absorbed more rapidly than the higher-energy photons. This energy dependence also allows the possibility of basis-material decomposition [5, 48, 66, 104, 142, 144]. For radiotherapy, an accurate image of attenuation coefficients at a higher treatment energy can be synthesized by combining component images separated at low diagnose energies. This accurate image ensures precise dose calculation, enhances visualization and thus segmentation of anatomy for radiotherapy treatment planning, and may lead to future improvements in reducing image artifacts from highly attenuating materials.

Dual-energy (DE) CT methods are the most common approaches for reconstructing two basis materials (soft-tissue and bone). However, DECT methods require either two scans or specialized scanner design, such as such as fast kVp-switching [48, 50], dual-source CT [112] and dual-layer detectors [39, 61]. We propose an alternative method in Chapter 4, which uses measurements from a single-energy scan acquired differential filtration, such as a X-ray split or bow-tie filter.

1.3 Motion Estimation for IGRT

Motion estimation provides movement information of the tumor and other critical organs to improve the effectiveness and efficiency of radiotherapy treatment. In general, 3D CT volumes are

available before treatment and can be used as reference images, while 2D projection radiographs at a limited range of angles are acquired during treatment. There are two classes of methods for estimating motion during treatment. One may reconstruct 3D target volumes using those radiographs from limited angles, and then carry out 3D-3D image domain registration between a reference image and target images. Such reconstructions are called tomosynthesis [41, 67, 124, 150]. One may also use 2D-3D image registration technology to estimation motion directly from those radiographs without reconstruction [77, 95, 106, 114, 115, 155, 158, 163, 164]. When the angular range of the CT scan is too small ($< 60^\circ$), the reconstructed images are corrupted with artifacts due to poor sampling, which affects the consequent image registration. A short angular range also affects the performance of 2D-3D image registration due to limited deformation information presented in the projection direction. We investigate the influence of angular range to 2D-3D image registration in Chapter 5. This investigation may help determine the tradeoffs associated with various parameters for position monitoring, such as projection arrangement, noise/dose, and temporal/spatial accuracy limits [11, 90, 109, 140].

1.4 Contributions and Outline

Image reconstruction and motion estimation are very important for IGRT. This thesis addresses several important issues related to these two core components.

We developed two new separable footprint (SF) projector methods that approximate the voxel footprint functions as 2D separable functions. Because of the separability of these footprint functions, calculating their integrals over a detector cell is greatly simplified and can be implemented efficiently. The SF-TR projector uses trapezoid functions in the transaxial direction and rectangular functions in the axial direction, whereas the SF-TT projector uses trapezoid functions in both directions. Simulations and experiments showed that both SF projector methods are more accurate

than the distance-driven (DD) projector [26], which is a current state-of-the-art method in the field. The SF-TT projector is more accurate than the SF-TR projector for rays associated with large cone angles. The SF-TR projector has similar computation speed with the DD projector and the SF-TT projector is about two times slower. To save computation and maintain relative accuracy, one may use the SF-TR projector for voxels that are near the X-ray source plane, where the cone angles are small and the rectangle approximation is reasonable, and use the SF-TT projector for other voxels associated with large cone angles.

We proposed a statistical penalized weighted least-squares (PWLS) method with edge-preserving regularizer for two-material decomposition from a single-energy CT scan acquired with differential X-ray filters, split and bow-tie filters. Differential filtration produced incident spectra variation among projection rays. We also proposed an optimization transfer method with separable quadratic surrogates to monotonically decrease the PWLS cost function which was non-convex and non-linear. We first reconstructed the bone-corrected FBP images using the Joseph and Spital method, and separated the soft and bone components by a threshold to initialize the iterative algorithm. The proposed PWLS method reconstructed soft tissue and bone components with 25% and 40% lower RMS errors respectively, and reconstructed density (the sum of soft tissue and bone) images and linear attenuation coefficient images at 511 Kev with 40% lower RMS errors than the JS-FBP method. The PWLS images produced less beam-hardening artifacts and noise than the JS-FBP method. The split and bow-tie filter led to similar results.

We investigated the fundamental accuracy limit of estimating local configuration of tumors using projection-to-volume alignment between a small spread of radiographs acquired during arc therapy and a reference CT volume. The projection-to-volume alignment procedure used a non-rigid model to describe motion in thorax area, a cost function consisting of a least-squared error metric and a simple regularizer that encourages local invertibility and a 4-level multi-resolution

scheme with a conjugate gradient method to optimize the cost function. The performed experiments demonstrated the potential accuracy of limited-angle projection-to-volume registration. Registration accuracy can be sensitive to angular center, tends to be larger along the projection direction, and tends to decrease away from the rotation center, whereas registration accuracy tends to be maintained at different noise levels and extents of deformation. This investigation indicates the potential of position monitoring of high contrast tumors during treatment using a small spread of projections without implanted markers.

This thesis is organized as follows. Chapter 2 briefly introduces the background on principles of X-ray CT, image registration and optimization transfer methods. Chapter 3 introduces our proposed 3D forward and back-projection methods for X-ray CT using separable footprints (SF). Chapter 4 presents the proposed PWLS method for two-material decomposition from a single scan. Chapter 5 describes our investigation into the influence of arc length on accuracy of motion estimation for projection-to-volume targeting during rotational therapy. Finally, we summarize our work and future research directions in Chapter 6.

CHAPTER 2

Background

2.1 Principles of X-Ray CT

X-ray CT produces images of the X-ray attenuation coefficient of the object or patient being scanned. A typical construction of a X-ray scanner involves a source and a detector array (see Fig. 2.1). The source and the detector array are fixed with respect to each other in space on a C-arm or a gantry and trace a path or orbit around the patient. The source produces incoherent X-ray radiation and detectors record the intensity of the radiation exiting the patient. As the source and the detector array scan the patient, each source position and detector element pair corresponds to a thin beam of radiation that passes through the patient and represents one measurement. The final image is generated from these measurements utilizing the basic principle that the internal structure of the body can be reconstructed from multiple X-ray projections.

2.1.1 General Measurement Model

We adopt the following general model to describe the measurement physics for X-ray CT. The detector measures X-ray photon emerging from the object at $M_0 \geq 1$ different incident spectra. Based on current technology, different incident spectra can be realized by either scanning with different X-ray spectra, such as fast kVp-switching [48, 50] and dual-source CT [112], or by dual-

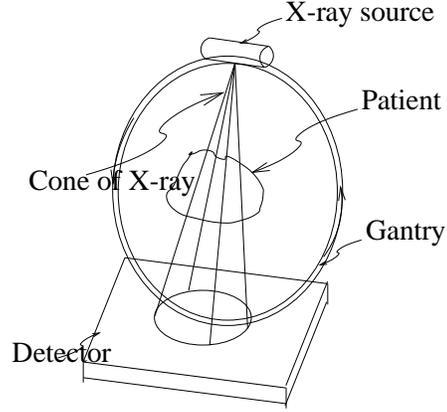


Figure 2.1: Schematic diagram of a X-ray scanner.

layer detectors [39, 61]. Let Y_{im} denote the measurement for the ray \mathcal{L}_{im} which is the i th ray for the m th energy scan, where $m = 1, \dots, M_0$, $i = 1, \dots, N_d$, and N_d is the number of rays. For notational simplicity we assume that the same number of rays are measured for each incident spectrum, but the physics model and methods presented in this dissertation can be easily generalized to cases where different incident spectra have different number of recorded rays. For a ray \mathcal{L}_{im} of infinitesimal width, the mean of the projection measurements could be expressed as:

$$\mathbb{E}_\mu[Y_{im}] = \bar{y}_{im} \triangleq \int I_{im}(\mathcal{E}) \exp\left(-\int_{\mathcal{L}_{im}} \mu(\vec{x}, \mathcal{E}) d\ell\right) d\mathcal{E} + r_{im}, \quad (2.1)$$

where $\mu(\vec{x}, \mathcal{E})$ denotes the 3D unknown *spatially- and energy-dependent* attenuation distribution, $\int_{\mathcal{L}_{im}} \cdot d\ell$ denotes the “line integral” function along line \mathcal{L}_{im} , and the incident X-ray intensity $I_{im}(\mathcal{E})$ incorporates the source spectrum and the detector gain. In reality, the measurements suffer from background signals such as Compton scatter, dark current and noise. The ensemble mean of those effects (for the ray \mathcal{L}_{im}) is denoted as r_{im} . We treat each $I_{im}(\mathcal{E})$ and r_{im} as known nonnegative quantities. In practice, $I_{im}(\mathcal{E})$ can be determined by careful calibration [122], and r_{im} are estimated by some preprocessing steps prior to iterative reconstruction [25, 156, 157].

From (2.1), we see that the CT measurements $\{Y_{im}\}_{i=1}^{N_d}$ indirectly correspond to the projections of an object’s attenuation coefficient $\mu(\vec{x}, \mathcal{E})$. The overall goal of X-ray CT image reconstruction

is to reconstruct $\mu(\vec{x}, \mathcal{E})$ from $\{Y_{im}\}_{i=1}^{N_d}$.

2.1.2 Cone-Beam Geometry

Any reconstruction method must account for the geometry of the imaging system, which is “hidden” in the line integrals $\{\int_{\mathcal{L}_{im}} \cdot d\ell\}$ in (2.1). This section elaborates on the geometry and coordinate systems.

Axial cone-beam projection space is characterized by three independent indices (s, t, β) and two distance parameters (D_{s0}, D_{0d}) , where β denotes the angle of the source point counter-clockwise from the y axis, (s, t) denote the detector coordinates, D_{s0} denotes the source to rotation center distance and D_{0d} denotes the isocenter to detector distance. (see Fig. 2.2). The axial cone-beam geometry is a special case of helical cone-beam geometry with zero helical pitch.

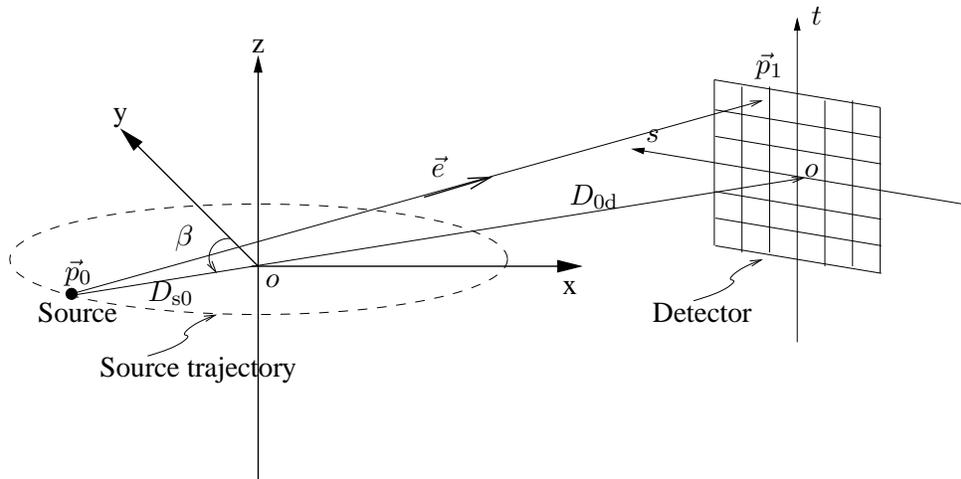


Figure 2.2: Axial cone-beam flat-detector geometry.

For simplicity of presentation, we focus on the flat-detector axial cone-beam geometry (see Fig. 2.2) hereafter. The formulas here and the methods in Chapter 3 generalize easily to arc detectors and helical geometries.

The source lies on points on a circle of radius D_{s0} centered at the rotation center on the $z = 0$

plane. The source position \vec{p}_0 can be parameterized as follows:

$$\vec{p}_0 = \begin{pmatrix} -D_{s0} \sin \beta \\ D_{s0} \cos \beta \\ 0 \end{pmatrix}. \quad (2.2)$$

For simplicity, we present the case of an ideal point source of X-rays.

Let (s, t) denote the local coordinates on the 2D detector plane, where the s -axis is perpendicular to the z -axis, and the t -axis is parallel to the z -axis. A point on the 2D detector can be expressed as

$$\vec{p}_1 = \begin{pmatrix} s \cos \beta + D_{0d} \sin \beta \\ s \sin \beta - D_{0d} \cos \beta \\ t \end{pmatrix}, \quad (2.3)$$

where $D_{0d} = D_{sd} - D_{s0}$ is the isocenter to detector distance. The direction vector of a ray from \vec{p}_0 to \vec{p}_1 can then be expressed as

$$\begin{aligned} \vec{e} &= \frac{\vec{p}_1 - \vec{p}_0}{\|\vec{p}_1 - \vec{p}_0\|} \\ &= \frac{1}{\sqrt{D_{sd}^2 + s^2 + t^2}} \begin{pmatrix} s \cos \beta + D_{sd} \sin \beta \\ s \sin \beta - D_{sd} \cos \beta \\ t \end{pmatrix} = \begin{pmatrix} \sin \varphi \cos \theta \\ -\cos \varphi \cos \theta \\ \sin \theta \end{pmatrix}, \end{aligned} \quad (2.4)$$

where

$$\gamma = \gamma(s) \triangleq \arctan\left(\frac{s}{D_{sd}}\right) \quad (2.5)$$

$$\varphi = \varphi(s, \beta) \triangleq \gamma(s) + \beta \quad (2.6)$$

$$\theta = \theta(s, t) \triangleq -\arctan\left(\frac{t}{\sqrt{s^2 + D_{sd}^2}}\right), \quad (2.7)$$

and φ and θ denote the azimuthal and polar angle of the ray from \vec{p}_0 to \vec{p}_1 respectively.

The **cone-beam projections** of a 3D object $\mu(\vec{x})$, where $\vec{x} = (x, y, z)$, are given by

$$p(s, t; \beta) = \int_{\mathcal{L}(s, t, \beta)} \mu(\vec{x}) \, d\ell, \quad (2.8)$$

where the integral is along the line segment:

$$\begin{aligned} \mathcal{L}(s, t, \beta) &= \{\vec{p}_0 + \ell \vec{e} : \ell \in [0, L_p]\} \\ L_p &\triangleq \sqrt{D_{sd}^2 + s^2 + t^2}. \end{aligned} \quad (2.9)$$

For a point $\vec{x} = (x, y, z)$ between the source and detector, the azimuthal and polar angles of the ray connecting the source and \vec{x} are

$$\varphi(\beta; x, y) = \beta + \arctan\left(\frac{\tau_p(\beta; x, y)}{d_s(\beta; x, y)}\right) \quad (2.10)$$

$$\theta(\beta; x, y, z) = -\arctan\left(\frac{z}{\sqrt{(\tau_p(\beta; x, y))^2 + (d_s(\beta; x, y))^2}}\right). \quad (2.11)$$

The projected s coordinate of \vec{x} is

$$\tau(\beta; x, y) = D_{sd} \frac{\tau_p(\beta; x, y)}{d_s(\beta; x, y)}, \quad (2.12)$$

where

$$\begin{aligned} \tau_p(\beta; x, y) &\triangleq x \cos \beta + y \sin \beta, \\ d_s(\beta; x, y) &\triangleq D_{s0} - \tau_{\perp}(\beta; x, y), \\ \tau_{\perp}(\beta; x, y) &\triangleq -x \sin \beta + y \cos \beta. \end{aligned} \quad (2.13)$$

The projected t coordinate is

$$t(\beta; x, y, z) = z \frac{D_{sd}}{d_s(\beta; x, y)}. \quad (2.14)$$

Given the points in the support of a 3D object $\mu(\vec{x})$, the support of its cone-beam projection function on the detector plane can be determined by the s and t coordinates of projections of those points. Note that a object close to the source has a bigger support than the same object close to the detector.

In practice, it is not feasible to evaluate the cone-beam projections (2.8) of an arbitrary object $\mu(\vec{x})$ on the fly. All the existing projectors make approximations about computing them based on the features of the geometry of imaging system and the object itself, as do our proposed SF projectors (Chapter 3).

2.1.3 Image Reconstruction Methods

The overall goal of x-ray CT image reconstruction is to reconstruct the underlying object being imaged from the projection measurements, particularly $\mu(\vec{x}, \mathcal{E})$ from $\{Y_{im}\}_{i=1}^{N_d}$ in transmission tomography. Image reconstruction methods or algorithms can be divided into two main categories: analytical methods, such as filtered back-projection (FBP) and its extension, Feldkamp (FDK) cone-beam approach [30], and iterative methods, such as statistical methods. Weighted least squares (WLS) and penalized likelihood (PL) reconstruction are statistical reconstruction methods.

Feldkamp (FDK) Cone-Beam Approach

Filtered backprojection (FBP) is the most common analytical reconstruction technique that is based on the Radon transform. The basic idea behind this method is to “smear” measured sinogram values back into the object space along the corresponding rays. This operation is called the *backprojection* operator that is the *adjoint* operator of the forward projection operation or Radon transform.

FDK approach [30] is the extension of FBP for axial cone-beam geometry. The basic idea of the FDK method is to filter the data measured along each row of the detector as if it were part of a 2D fan-beam acquisition, and then to perform 3D backprojection. With this approximation, the central slice is reconstructed exactly as it would be by conventional 2D fan-beam methods, whereas the non-central slices have artifacts that increase with distance from the central slice.

Conventional analytical methods, including FBP and FDK, assume *single monoenergetic* property of the X-ray spectrum:

$$I_{im}(\mathcal{E}) = I_i(\mathcal{E}) \delta(\mathcal{E} - \mathcal{E}_0), \quad (2.15)$$

where \mathcal{E}_0 denotes the single energy, and $\delta(\cdot)$ denotes a Dirac impulse. We ignore the index m from now on since $M_0 = 1$ for single energy measurements. Under this assumption, the mean intensities in (2.1) simplifies to

$$\bar{y}_i = I_i(\mathcal{E}_0) \exp\left(-\int_{\mathcal{L}_i} \mu(\vec{x}, \mathcal{E}_0) d\ell\right) + r_i. \quad (2.16)$$

The estimated line integrals $\ell_i \triangleq \int_{\mathcal{L}_i} \mu(\vec{x}, \mathcal{E}_0) d\ell$ can be obtained by taking the logarithm of the measurements:

$$\hat{\ell}_i \triangleq \log\left(\frac{I_i}{y_i - r_i}\right) \approx \int_{\mathcal{L}_i} \mu(\vec{x}, \mathcal{E}_0) d\ell. \quad (2.17)$$

For cone-beam geometry, the goal of FDK image reconstruction is to estimate $\mu(\vec{x})$ at energy \mathcal{E}_0 , from measured cone-beam projections $\hat{p}(s, t; \beta)$ (akin to $\hat{\ell}_i$) that are obtained from (noisy) samples of projection data using (2.17).

The FDK algorithm for a flat-detector axial cone-beam scanner consists of the following steps [30].

- Step 1. Compute weighted projections

$$\tilde{p}(s, t; \beta) = w_1(s, t) \hat{p}(s, t; \beta), \quad w_1(s, t) = \frac{D_{s0} \sqrt{1 + (t/D_{sd})^2}}{\sqrt{D_{sd}^2 + s^2 + t^2}}. \quad (2.18)$$

- Step 2. Filter each row of those projections using a ramp filter $h(s)$:

$$\check{p}(s, t; \beta) = \tilde{p}(s, t; \beta) * h(s). \quad (2.19)$$

The frequency response of ramp filter $h(s)$ is $H(\nu) = |\nu|$.

- Step 3. Perform weighted cone-beam backprojection of those filtered projections:

$$\begin{aligned} \hat{\mu}(x, y, z) &= \frac{1}{2} \int_0^{2\pi} w_2(x, y, \beta) \check{p}\left(\frac{D_{sd}}{D_{s0} - y_\beta} x_\beta, \frac{D_{sd}}{D_{s0} - y_\beta} z; \beta\right) d\beta, \\ w_2(x, y, \beta) &= \frac{D_{sd}^2}{(D_{s0} - y_\beta)^2}, \end{aligned} \quad (2.20)$$

where the factor $\frac{D_{sd}}{D_{s0} - y_\beta}$ is a magnification factor, and the rotated coordinates (x_β, y_β) is defined as $x_\beta \triangleq x \cos \beta + y \sin \beta$ and $y_\beta \triangleq -x \sin \beta + y \cos \beta$.

In reality, the ideal ramp filter cannot be implemented on discrete data and must be set zero beyond certain cutoff frequency. Also note that ramp filter amplifies noise. Therefore, an optimal reconstruction is not possible for FDK approach from noisy projections.

Statistical Image Reconstruction

Statistical image reconstruction methods are based on measurement statistics and physics models and offer numerous advantages, such as the potential for improved bias-variance performance and providing quantitatively accurate CT values.

For the case of normal clinical exposures, the X-ray CT measurements are often modeled as the sum of a *Poisson* distribution representing photon-counting statistics (2.1) and a zero-mean normal distribution representing additive electronic noise:

$$Y_{im} \sim \text{Poisson}\{\bar{y}_{im}\} + \text{N}(0, \sigma_e^2), \quad (2.21)$$

where σ_e denotes the standard deviation of electronic noise. Because (2.21) does not leads to a tractable likelihood function, an approximate shifted Poisson likelihood function that matches the

first and second moments is often used [131, 157]:

$$Y_{im} + \sigma_e \sim \text{Poisson}\{\bar{y}_{im} + \sigma_e^2\}. \quad (2.22)$$

Absorbing the electronic noise into r_{im} , we obtain the following measurement statistics

$$Y_{im} \sim \text{Poisson}\left\{\int I_{im}(\mathcal{E}) \exp\left(-\int_{\mathcal{L}_{im}} \mu(\vec{x}, \mathcal{E}) d\ell\right) d\mathcal{E} + r_{im}\right\}. \quad (2.23)$$

We discretize the continuous attenuation function $\mu(\vec{x}, \mathcal{E})$ and parameterize it using basis functions that are separable in both the spatial and energy dimensions [5, 24, 48, 142, 144] as follows,

$$\mu(\vec{x}, \mathcal{E}) = \sum_{l=1}^{L_0} \sum_{j=1}^{N_p} \beta_l(\mathcal{E}) b_j(\vec{x}) x_{lj} \quad (2.24)$$

where $\beta_l(\mathcal{E})$ is the energy-dependent mass attenuation coefficient of the l th material type (units cm^2/g), $\{b_j(\vec{x})\}$ are unitless spatial basis functions, and x_{lj} is the unknown density of the l th material type in the j th voxel (units g/cm^3). Now the reconstruction goal becomes to estimate the object $\{x_{lj}\}_{j=1}^{N_p}$ from $\{Y_{im}\}_{i=1}^{N_d}$.

For simplicity, we use the *monoenergetic measurement* model given in (2.16) to demonstrate the ideas behind statistical image reconstruction. In the case of a single monoenergetic source, we usually assume

$$\mu(\vec{x}, \mathcal{E}_0) = \sum_{j=1}^{N_p} \beta(\mathcal{E}_0) b_j(\vec{x}) x_j \quad (2.25)$$

for some spatial basis functions $b_j(\cdot)$, such as cubic voxels. Substituting into (2.16) yields

$$\bar{y}_i = I_i(\mathcal{E}_0) e^{-\beta(\mathcal{E}_0)[\mathbf{A}\mathbf{x}]_i} + r_i, \quad (2.26)$$

where \mathbf{A} is the system matrix with entries

$$a_{ij} \triangleq \int_{\mathcal{L}_i} b_j(\vec{x}) d\ell. \quad (2.27)$$

Then (2.23) simplifies to

$$Y_i \sim \text{Poisson}\{I_i(\mathcal{E}_0) e^{-\beta(\mathcal{E}_0)[\mathbf{A}\mathbf{x}]_i} + r_i\} \quad (2.28)$$

The reconstruction goal is to estimate the vector of linear attenuation coefficient values $\mathbf{x} = (x_1, \dots, x_{N_p})$ from $\{Y_i\}_{i=1}^{N_d}$.

Penalized-Likelihood (PL) Reconstruction

For the mono-energetic and single material object model (2.28), the negative log-likelihood has the form

$$\begin{aligned} -L(\mathbf{x}) &= \sum_{i=1}^{N_d} \{Y_i \log(\bar{Y}_i(\mathbf{x})) - \bar{Y}_i(\mathbf{x})\} \\ &= \sum_{i=1}^{N_d} \{Y_i \log(I_i(\mathcal{E}_0) e^{-\beta(\mathcal{E}_0)[\mathbf{A}\mathbf{x}]_i} + r_i) - (I_i(\mathcal{E}_0) e^{-\beta(\mathcal{E}_0)[\mathbf{A}\mathbf{x}]_i} + r_i)\} \end{aligned} \quad (2.29)$$

where $\bar{Y}_i(\mathbf{x}) \triangleq E[Y_i|\mathbf{x}]$ is the mean of the measurement data along path L_i (We simplify the subscript $1i$ to i). Objective functions based solely on the negative log-likelihood (2.29) perform poorly due to the ill-conditioned nature of tomographic reconstruction. Unregularized methods provide increasingly noisy images with iteration [132]. To remedy this problem, we add a roughness penalty term or “prior” [36] to the negative log-likelihood, resulting in the following *penalized-likelihood (PL)* cost function:

$$\Psi(\mathbf{x}) = -L(\mathbf{x}) + \beta R(\mathbf{x}), \quad (2.30)$$

where the (pairwise) penalty term has the following form

$$R(\mathbf{x}) = \sum_{j=1}^{N_p} \sum_{k \in \mathcal{N}_j} \psi(x_j - x_k), \quad (2.31)$$

where $\psi(\cdot)$ is an edge-preserving potential function [42, 70] and \mathcal{N}_j is some neighborhood of voxel j .

The effect of the penalty term is to discourage disparities between neighboring voxel values, while the effect of the first term in (2.30) is to encourage agreement with the measured data. These are usually conflicting goals, and the smoothing parameter β controls the trade-off between the two.

Penalized Weighted Least-Squares (PWLS) Reconstruction

The Poisson log-likelihood in (2.29) is based on the statistical properties of the tomographic reconstruction problem. However, because Poisson negative log-likelihood (2.29) is non-convex and non-quadratic, the minimization algorithms required can be complex. To simplify, one can apply a second-order Taylor expansion to the Poisson log-likelihood in (2.29) [31, 32, 125]. This quadratic approximation leads to weighted least squares (WLS) likelihood function:

$$-L(\mathbf{x}) = \sum_{i=1}^{N_d} w_i \frac{1}{2} (-\log(Y_i) - [\mathbf{A}\mathbf{x}]_i)^2, \quad (2.32)$$

where w_i values are statistical weighting factors that depend on the model for the measurement statistics. For the case of regular Poisson likelihood, $w_i \approx \frac{(\bar{Y}_i(\mathbf{x}) - r_i)^2}{Y_i(\mathbf{x})}$. For large means, *i.e.*, high incident intensities $I_i(\mathcal{E}_0)$, the Poisson distribution is approximately Gaussian by the Central Limit Theorem [13]. Gaussian approximation could also lead to the WLS likelihood function (2.32). Plugging (2.32) into (2.30) yields penalized weighted least squares (PWLS) cost function. In practice, one often uses simply $w_i = Y_i$.

Gradient-based iterative optimization algorithms are often applied to minimize (2.30) subject to certain object constraints such as non-negativity:

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x} \geq \mathbf{0}} \Psi(\mathbf{x}). \quad (2.33)$$

The column gradient of the cost function in (2.30) has the follow matrix form:

$$\nabla \Psi(\mathbf{x}) = -\mathbf{A}' \text{diag}\{Y_i/\bar{Y}_i(\mathbf{x}) - 1\} \nabla_{\mathbf{x}} \bar{\mathbf{Y}}(\mathbf{x}) + \beta \nabla \mathbf{R}. \quad (2.34)$$

From (2.28) and (2.34), we can see that one forward projection and one backprojection are involved in each iteration. Therefore, fast forward and back-projectors are essential in computation time reduction. This is the main motivation for our work on the development of fast and accurate 3D forward and back projectors in Chapter 3.

Iterative Reconstruction of A Region of Interest (ROI)

In many cases, the region of interest (ROI) needed for diagnosis is much smaller than the field of view (FOV) that covers the whole irradiated volume. One such application is cardiac cone-beam CT, where the heart defines the ROI. When the ROI is known in advance, *e.g.*, from a scout scan, a ROI reconstruction could save resources in terms of computation time and memory use.

For exact and approximative analytical reconstruction methods based on FBP, ROI reconstruction is possible without any extra effort. However, iterative reconstruction methods use the forward projection and the measurement to calculate an update that improves the image. This implies the requirement that the entire FOV has to be reconstructed that contributed to the absorption. Ziegler *et al.* [168] proposed the following solution to iterative reconstruction of a ROI.

Let $\mathbf{p} = \mathbf{A}\mathbf{x}$ denote the forward projection of the density vector $\mathbf{x} = (x_1, \dots, x_{N_p})$. The general idea of iterative ROI reconstruction consists of five consecutive steps.

1. Iterative reconstruction of the whole FOV, yielding an initial estimate $\hat{\mathbf{x}} = (\hat{x}_1, \dots, \hat{x}_{N_p})$ of \mathbf{x} .
2. Define $\hat{\mathbf{x}}_{\text{mask}} = (\hat{x}_1 m_1, \dots, \hat{x}_{N_p} m_{N_p})$ where $\mathbf{m} = (m_1, \dots, m_{N_p})$ with $0 \leq m_j \leq 1 (j = 1, \dots, N_p)$ is a mask vector setting the density $\hat{\mathbf{x}}$ of the object, inside the ROI to zero and provides a smooth transition from the ROI to the remaining voxels.
3. Compute $\mathbf{p}_{\text{out}} = \mathbf{A}\hat{\mathbf{x}}_{\text{mask}}$ which is the forward projection of the masked density image $\hat{\mathbf{x}}_{\text{mask}}$.
4. Obtain the projection of ROI, $\mathbf{p}_{\text{roi}} = \mathbf{p} - \mathbf{p}_{\text{out}}$.

5. Iterative reconstruction of the ROI only with \mathbf{p}_{roi} . Due to the transition zone, the region of this reconstruction needs to be extended slightly from the predetermined ROI.

The re-projection through the density image with the removed ROI has to model the geometry of the initial measurement properly, because the calculation of the difference between the re-projection and the measurement is not consistent otherwise. This can only be achieved by accurate 3D projectors.

2.1.4 Beam-Hardening Correction

A X-ray beam is composed of photons with a broad energy spectrum. Beam hardening is the phenomenon that the mean energy of a X-ray beam increases as it penetrates through an object because the lower energy photons are absorbed more readily than the higher energy photons. If this beam-hardening effect is ignored, two types of artifacts will appear in the reconstructed images: cupping and streaks. The cupping artifacts are caused by the fact that the X-rays passing through the middle of an object are hardened more than the ones passing through the edges because they are passing through more material. As the beam becomes harder, the rate at which it is attenuated decreases. The resultant profiles through the reconstructed image of a uniform cylindrical object display a cupped shape compared to the ideal profiles without beam hardening. Streak artifacts are typically present between two dense objects, such as two bony regions of the body. These artifacts are caused by the fact the portion of the beam passing through one of the dense objects at certain X-ray source positions is hardened less than when it passes through both objects at other source positions.

Current beam-hardening correction methods can be divided into single-energy [3, 28, 29, 57, 69, 135] and dual-energy [5, 35, 37, 48, 104, 144] correction techniques. Most methods are based on classifying the object materials into two categories: soft tissue and bone. “Soft tissue” means

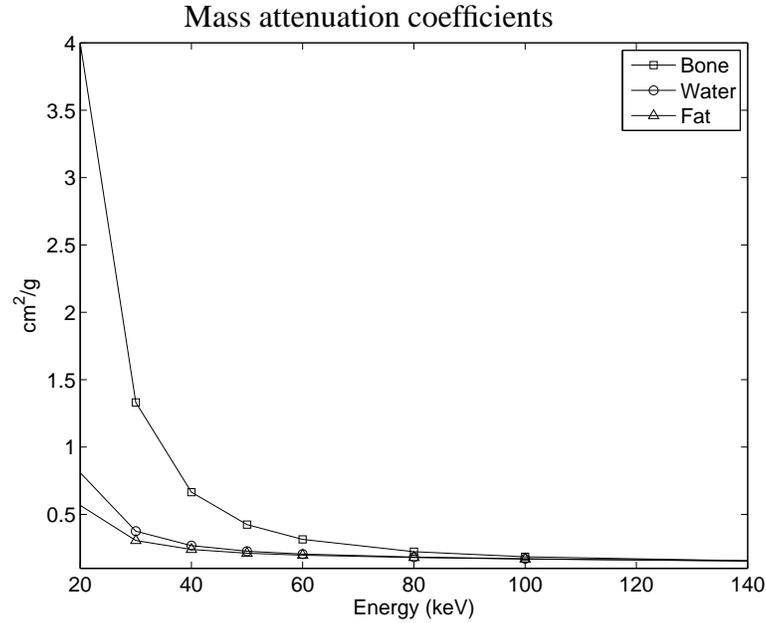


Figure 2.3: Energy-dependent mass attenuation coefficients $\beta_l(\mathcal{E})$ for bone, fat and water.

any tissue whose mass attenuation coefficient is sufficiently close to that of water. Fig. 2.3 shows typical mass attenuation coefficients for water, fat and bone. The curves of water and fat almost overlap. Water and bone corrections are the two most commonly used single-energy techniques. The water correction technique assumes the object consists of only water equivalent materials and corrects the measurements prior to reconstruction [43, 133]. This often leads to suboptimal correction, especially for non-homogeneous objects with high density areas, like bones. Bone correction techniques usually perform water correction first and then perform additional correction for bone [57]. Dual-energy (DE) CT imaging is considered as a more theoretically elegant approach to eliminate beam hardening artifact [28]. Typically, this technique uses two sets of measurements, one at a lower energy and the other at a higher energy, to estimate two basis material (*e.g.*, soft-tissue and bone) images. An attenuation coefficient image can in principle be presented at any energy, free from beam hardening artifacts. Typical DECT methods require either two data acquisitions or specialized scanner designs, such as fast kVp-switching, dual-source CT and dual-layer

detectors. Several methods have been developed to reduce beam-hardening artifacts from reconstructed soft tissue and bone images from one CT scan and a single tube voltage setting through special filtration [121, 123, 146]. We summarize the water and bone correction techniques here, and describe the DECT method in Chapter 4.

Assume two basis materials, soft tissue and bone, are present, *i.e.*,

$$\mu(\vec{x}, \mathcal{E}) = \beta_S(\mathcal{E}) \rho_S(\vec{x}) + \beta_B(\mathcal{E}) \rho_B(\vec{x}), \quad (2.35)$$

where $\beta_S(\mathcal{E})$ and $\beta_B(\mathcal{E})$ are the energy-dependent mass attenuation coefficients of soft tissue (water) and bone, and $\rho_S(\vec{x})$ and $\rho_B(\vec{x})$ are the density of soft tissue and bone at spatial location \vec{x} respectively. Water and bone correction require only one single scan, *i.e.*, $M_0 = 1$, and (2.1) becomes

$$E_\mu[Y_i] = \bar{y}_i \triangleq \int I_i(\mathcal{E}) \exp\left(-\int_{\mathcal{L}_i} \mu(\vec{x}, \mathcal{E}) d\ell\right) d\mathcal{E} + r_i. \quad (2.36)$$

Combining the measurement model (2.36) and the object model (2.35), the mean of measurements can be represented as follows

$$\bar{y}_i = I_i e^{-f_i(T_{S,i}, T_{B,i})} + r_i, \quad (2.37)$$

where

$$f_i(T_{S,i}, T_{B,i}) \triangleq -\log \left\{ \int \frac{I_i(\mathcal{E})}{I_i} e^{-(\beta_S(\mathcal{E}) T_{S,i} + \beta_B(\mathcal{E}) T_{B,i})} d\mathcal{E} \right\} \quad (2.38)$$

$$I_i \triangleq \int I_i(\mathcal{E}) d\mathcal{E}. \quad (2.39)$$

The nonlinear functions $f_i(T_{S,i}, T_{B,i})$ characterize the beam hardening caused by polychromatic source spectra. The total intensity I_i of the incident spectrum for the i th ray is defined in (2.39).

The line integral functions $T_{S,i}$ and $T_{B,i}$ are defined as

$$T_{S,i} \triangleq \int_{\mathcal{L}_i} \rho_S(\vec{x}) d\ell \quad (2.40)$$

$$T_{B,i} \triangleq \int_{\mathcal{L}_i} \rho_B(\vec{x}) d\ell. \quad (2.41)$$

Given measurements \bar{y}_i , the $f_i(T_{S,i}, T_{B,i})$ values are usually estimated by inverting (2.37) as follows:

$$\hat{f}_i \triangleq -\log\left(\frac{Y_i - r_i}{I_i}\right). \quad (2.42)$$

Water Correction

Most clinical CT scanners use the water correction method to eliminate the cupping artifacts in soft tissues. This method assumes only soft tissues are present, *i.e.*, (2.35) becomes

$$\mu(\vec{x}, \mathcal{E}) = \beta_s(\mathcal{E}) \rho_s(\vec{x}). \quad (2.43)$$

Ignore the path-dependent subscript i hereafter and define function $f_s(T_e)$ as

$$f_s(T_e) \triangleq f(T_e, 0), \quad (2.44)$$

where T_e is the "effective water path length".

Fig. 2.3 shows the mass attenuation coefficient of water $\beta_s(\mathcal{E})$ from NIST web pages¹. Given the incident spectrum for each path \mathcal{L}_i , *i.e.*, assuming $I_i(\mathcal{E})$ is known, one can exactly evaluate $f_s(T_e)$ for various known thicknesses of T_e using (2.38). Without the knowledge of the spectrum, one can scan water-only calibration phantoms with a known shape (usually a cylinder) and calculate the line integral function T_e along different paths. Then $f_s(T_e)$ is known as a function of T_e . With $f_s(T_e)$ tabulated, one can estimate \hat{T}_e using interpolation from

$$\hat{T}_e \triangleq f_s^{-1}(\hat{f}), \quad (2.45)$$

assuming $\hat{f} \approx f_s(T_e)$ for some T_e . One also can approximate f_s^{-1} using a polynomial function. An empirical method [133] is proposed to perform water correction, requiring neither knowledge of

¹ <http://physics.nist.gov/PhysRefData/XrayMassCoef/ComTab/water.html>

the spectrum or of the attenuation coefficients, nor the exact knowledge of the calibration phantom size and position.

One can apply the regular FDK method mentioned in Section 2.1.3 to the estimated line integrals \hat{T}_e to obtain the estimated density $\hat{\rho}_e(\vec{x})$, and then multiply $\hat{\rho}_e(\vec{x})$ by the mass attenuation coefficient of soft tissue $\beta_s(\mathcal{E}_0)$ at energy \mathcal{E}_0 to yield the attenuation image $\mu(\vec{x}, \mathcal{E}_0)$. Comparing equations (2.17) and (2.45), one can see that water-corrected reconstruction partially takes into account the polyenergetic property of the incident spectrum by estimating \hat{T}_e by inverting the beam-hardening function f_s .

Bone Correction

Since the human body does not consist of only soft tissues, the reconstructed image using the water-corrected method will be biased. Therefore, further correction for high-density materials, mainly bones, is often required. We review the well-known bone correction method proposed by Joseph and Spital [57], which we call the JS-FBP Method in this thesis.

Mass attenuation coefficients of soft tissue $\beta_s(\mathcal{E})$ and bone $\beta_B(\mathcal{E})$ are energy-dependent. Define λ as the ratio

$$\lambda(\mathcal{E}) = \beta_B(\mathcal{E}) / \beta_s(\mathcal{E}). \quad (2.46)$$

If λ were completely independent of energy, then the soft tissue correction function (2.45) would lead to artifact-free reconstructions, where the output would be an “effective density”

$$\rho_e(\vec{x}) = \rho_s(\vec{x}) + \lambda\rho_B(\vec{x}). \quad (2.47)$$

The natural physical interpretation of this equation is that 1 g/cm³ of bone is effectively λ times more dense than 1 g/cm³ of soft tissue. Simply using the water correction function (2.45) will

generate a series of “effective water path lengths”:

$$\hat{T}_e = \hat{T}_S + \lambda_i \hat{T}_B. \quad (2.48)$$

The main problem is that λ_i is both path and energy dependent. This leads to inconsistent projection data that causes artifacts. In fact, the exact density that appears for line \mathcal{L}_i will be a function of both the thickness of water equivalent substance T_S and bone T_B in the path.

The JS-FBP method can be described as follows:

1. Perform water-corrected reconstruction to obtain an “effective density” image $\hat{\rho}_e(\vec{x})$ from \hat{T}_e using the method mentioned in Section 2.1.4.
2. Segment the regions of $\hat{\rho}_e(\vec{x})$ that exceed a given threshold ρ_1 as bone regions.

This step assumes that all significant densities encountered in human body are basically “bone like”. It is stated in [57] that even without polychromatic effects, the combined effect of the initial water-corrected reconstruction and the finite beam width is to enlarge the bones in such a way that many values of $\hat{\rho}_e(\vec{x})$ are seriously overestimated. Setting $\rho_1 = 1.4\text{g/cm}^3$, 40% above soft tissue density was found to be able to provide satisfactory results.

3. Forward project the estimated bone image and divide it by λ_0 to obtain \hat{T}_B , the estimate of bone density integral.

One can inspect actual λ_i values in (2.48) from a few projections and select a typical value to be λ_0 . λ_i is a rather slowly varying function of T_B , and the value of λ_i obtained are rather insensitive to λ_0 [57].

4. Modify the water equivalent length \hat{T}_e according to

$$T_0 = \hat{T}_e + (\lambda_0 - \lambda_i) \hat{T}_B. \quad (2.49)$$

$\lambda_i(T_S, T_B)$ is the function that satisfies $f(T_S, T_B) = f_S(T_S + \lambda_i(T_S, T_B)T_B, 0)$. One can solve this equation numerically to obtain the values of λ_i as a function of T_S and T_B . It was shown [57] that if T_B is small enough,

$$\lambda_i \approx A - BT_B, \quad (2.50)$$

where

$$\begin{aligned} A &= \frac{\langle \beta_B(\mathcal{E}) \rangle}{\langle \beta_S(\mathcal{E}) \rangle} \\ B &= \frac{T_B}{2 \langle \beta_S(\mathcal{E}) \rangle} \left\langle \left(T_B - T_S \frac{\langle T_B \rangle}{\langle T_S \rangle} \right)^2 \right\rangle, \end{aligned} \quad (2.51)$$

and $\langle \cdot \rangle$ indicates averaging over the energy spectrum of the transmitted beam. It is practical to estimate the detected spectrum by averaging the exiting spectra of various rays through the object.

5. Apply analytical reconstruction method, such as FBP or FDK, or an iterative reconstruction method, to the modified line integrals T_0 to obtain the final corrected density image.

2.2 2D-3D Image Registration

Image registration is the process of overlaying two images of the same object taken at different times, from different viewpoints, and/or by different imaging technologies [169]. Mathematically, given a continuous-space reference image $f^{\text{ref}}(\vec{r})$ (where $\vec{r} \triangleq (x, y, z)$) and a target image $f^{\text{tar}}(\vec{r})$, the task of image registration is to find a geometric transformation $\mathcal{T} : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ such that $\mathcal{T} f^{\text{ref}}(\vec{r})$ is similar to $f^{\text{tar}}(\vec{r})$. Applications of image registration in medicine include combining images of the same subject from different modalities (CT, PET, MRI, etc.), aligning temporal sequences of images to compensate for motion of the subject between scans, image guidance during interventions and aligning images from multiple subjects in cohort studies [44].

However, for some applications a target image whose dimension is the same as the reference image is not available. For example, 3D CT images are used for clinical diagnosis and treatment planning, while 2D X-ray fluoroscopy or electronic portal images are acquired during treatment. These 2D images are real-time, but they lack the spatial information contained in 3D CT images. There are also some important anatomical features which are not visualized well in these 2D images, but can be observed in 3D CT images. 2D-3D image registration is a technology to achieve this kind of registration [76, 107].

2D-3D image registration is the process of aligning a 3D reference image $f^{\text{ref}}(\vec{r})$ to a series of target 2D projection images \mathbf{y}_n for $n = 1, \dots, N$ where N is the number of projection views. It assumes that the reference image can be converted to simulated projection images, such as digitally rendered radiographs (DRR) by system operators $\mathcal{A}_{\varphi_n} : \mathcal{L}_2(\mathbb{R}) \rightarrow \ell_2$ which captures the physics and geometries of the imaging modality used to produce the target 2D images, *i.e.*, $\mathcal{A}_{\varphi_n} f^{\text{ref}}(\vec{r})$ corresponds to \mathbf{y}_n which is the projection image at angle φ_n . A transformation \mathcal{T} is found by comparing these simulated projection images $\{\mathcal{A}_{\varphi_n} f^{\text{ref}}(\vec{r})\}_{n=1}^N$ with the target images $\{\mathbf{y}_n\}_{n=1}^N$. 2D-3D image registration has been widely used for patient set up estimation in radiotherapy system [4, 63] and motion estimation for targeting treatment [77, 95, 106, 114, 115, 155, 158, 163, 164].

In general, 2D-3D image registration is posed as an optimization problem as follows

$$\hat{\mathcal{T}} = \arg \min_{\mathcal{T} \in \mathcal{S}_{\text{trans}}} \mathbf{S}_{\text{im}}(\{\mathbf{y}_n\}, \{\mathcal{A}_{\varphi_n} \mathcal{T} f^{\text{ref}}(\vec{r})\}) + R(\mathcal{T}), \quad (2.52)$$

where $\mathcal{S}_{\text{trans}}$ is the set of desired geometric transformations, such as nonrigid transformations, $\mathbf{S}_{\text{im}}(\cdot, \cdot)$ is an intensity-based similarity measure, and $R(\mathcal{T})$ is a penalty function that discourages undesired \mathcal{T} . Except the three essential parts of image domain registration, similarity measure, geometric transformation and penalty function, 2D-3D registration has another important component, the system operator. We briefly review the first three ingredients in the following sections

and discuss the system operator in Chapter 3.

2.2.1 Similarity Measures

There are two types of differences between target projection images and DRRs of the reference image: one is associated with the changes in the imaged object and the other is due to differences in image formation. Image registration is aimed to express the first type of difference geometrically. Prior knowledge about this difference determines the choice of geometric transformation. For example, nonrigid transformation for soft-tissue deformation and rigid transformation for movement in bony areas. The second type of difference determines the choice of similarity measure. The sum of squared difference (SSD) metric can meaningfully compare images from the same modality,

$$\mathbf{S}_{\text{im}}(\{\mathbf{y}_n\}, \{\mathcal{A}_{\varphi_n}\}) = \frac{1}{2} \sum_{n=1}^N \|\mathbf{y}_n - \mathcal{A}_{\varphi_n} \mathcal{T} f^{\text{ref}}(\vec{\mathbf{r}})\|^2. \quad (2.53)$$

To use this metric, the X-ray energies should be the same for imaging the reference CT volume and the projection images, and extra efforts may be needed to correct the imaging artifacts caused by Compton scatter, beam hardening and presence of the radiotherapy table during treatment [164]. The SSD measure is the focus of this thesis.

Correlation-based metric compares images in terms of the correlation coefficient of their intensities [76, 164]. Images taken at different X-ray energy spectra are expected to have linearly related intensities. Mutual information criterion is useful for inter-modality registration. It assumes that the co-occurrence of the most probable values in the two images is maximized at registration [62, 76].

2.2.2 Nonrigid Deformation Model

Nonrigid transformation is suitable to describe changes in soft-tissue, such as the liver, heart and lung. We use a parametrized deformation model based on cubic B-splines [151]. Denote the

operator $\mathcal{T}(\vec{\mathbf{r}}; \boldsymbol{\alpha}) : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ that represents a nonrigid transformation described by unknown parameters $\boldsymbol{\alpha} \triangleq (\boldsymbol{\alpha}_X, \boldsymbol{\alpha}_Y, \boldsymbol{\alpha}_Z) \in \mathbb{R}^{N_\alpha}$ as

$$\mathcal{T}(\vec{\mathbf{r}}; \boldsymbol{\alpha}) = \vec{\mathbf{r}} + \mathcal{D}(\vec{\mathbf{r}}; \boldsymbol{\alpha}), \quad (2.54)$$

where $\vec{\mathbf{r}} \triangleq (x, y, z)$, x , y and z denote the left-right (LR), posterior-anterior (AP), superior-inferior (SI) direction respectively, $\mathcal{D}(\vec{\mathbf{r}}; \boldsymbol{\alpha}) = (\mathcal{D}_X(\vec{\mathbf{r}}; \boldsymbol{\alpha}), \mathcal{D}_Y(\vec{\mathbf{r}}; \boldsymbol{\alpha}), \mathcal{D}_Z(\vec{\mathbf{r}}; \boldsymbol{\alpha}))$ is the deformation map operator modeled by a tensor product of cubic B-splines β_3 as

$$\mathcal{D}_C(\vec{\mathbf{r}}; \boldsymbol{\alpha}) = \sum_{l=1}^{L_C} \alpha_{Cl} \beta_3\left(\frac{x}{\Delta_X} - m_{Xl}\right) \beta_3\left(\frac{y}{\Delta_Y} - m_{Yl}\right) \beta_3\left(\frac{z}{\Delta_Z} - m_{Zl}\right), \quad (2.55)$$

where $C \in \{X, Y, Z\}$, Δ_c is the knot spacing in the C direction, and $(m_{Xl}, m_{Yl}, m_{Zl}), l = 1, \dots, L_C$ are the knot coordinates.

Rather than operate on a continuous reference image $f^{\text{ref}}(\vec{\mathbf{r}})$, we represent it as a discretized object by a common basis function, such as cubic B-splines, as follows:

$$f^{\text{ref}}(x, y, z) = \sum_{k=1}^{N_p} u_k w_k(x, y, z), \quad (2.56)$$

where

$$w_k(x, y, z) = \beta_3(x - x_k) \beta_3(y - y_k) \beta_3(z - z_k), \quad (2.57)$$

$\mathbf{u} = (u_1, \dots, u_{N_p})$ is the vector of basis coefficients computed from the sample values of $f^{\text{ref}}(\vec{\mathbf{r}})$ by recursive digital filtering [151], and integer coordinates (x_k, y_k, z_k) denote centers of basis functions. We denote the reference image coefficient vector as \mathbf{f}^{ref} whose j -th component is

$$f_j^{\text{ref}} \triangleq f^{\text{ref}}(x_j, y_j, z_j). \quad (2.58)$$

Apply the same basis expansion model (2.56) to a target image $f^{\text{tar}}(\vec{\mathbf{r}})$ to obtain a target image coefficient vector \mathbf{f}^{tar} .

We next represent the transformation and deformation map operator in matrix-vector notation [53]. Define $\boldsymbol{\nu}_X$, $\boldsymbol{\nu}_Y$, and $\boldsymbol{\nu}_Z$, all in \mathbb{R}^{N_p} , as the vectors whose j -th components are x_j , y_j , and z_j respectively, and $\boldsymbol{\nu} \triangleq (\boldsymbol{\nu}_X, \boldsymbol{\nu}_Y, \boldsymbol{\nu}_Z)$. Define the matrices \mathbf{B}_X , \mathbf{B}_Y , and \mathbf{B}_Z to have entries

$$[\mathbf{B}_C]_{jl} \triangleq \beta_3 \left(\frac{x_j}{\Delta_X} - m_{Xl} \right) \beta_3 \left(\frac{y_j}{\Delta_Y} - m_{Yl} \right) \beta_3 \left(\frac{z_j}{\Delta_Z} - m_{Zl} \right), \quad C \in \{X, Y, Z\}, \quad (2.59)$$

and

$$\mathbf{B} \triangleq \begin{bmatrix} \mathbf{B}_X & & \\ & \mathbf{B}_Y & \\ & & \mathbf{B}_Z \end{bmatrix}. \quad (2.60)$$

Define the deformation map vector $\mathbf{D}(\boldsymbol{\alpha}) : \mathbb{R}^{3N_p} \rightarrow \mathbb{R}^{3N_p}$

$$\mathbf{D}(\boldsymbol{\alpha}) \triangleq (\mathbf{D}_X(\boldsymbol{\alpha}), \mathbf{D}_Y(\boldsymbol{\alpha}), \mathbf{D}_Z(\boldsymbol{\alpha})) = \mathbf{B}\boldsymbol{\alpha} = (\mathbf{B}_X\boldsymbol{\alpha}_X, \mathbf{B}_Y\boldsymbol{\alpha}_Y, \mathbf{B}_Z\boldsymbol{\alpha}_Z). \quad (2.61)$$

Assuming that the target image \mathbf{f}^{tar} is deformable from \mathbf{f}^{ref} , the geometric correspondence between them is

$$\mathbf{f}^{\text{tar}} = \mathbf{T}(\boldsymbol{\alpha})\mathbf{f}^{\text{ref}}, \quad (2.62)$$

where $\mathbf{T}(\boldsymbol{\alpha}) : \mathbb{R}^{N_p} \rightarrow \mathbb{R}^{N_p}$ denotes the operator that maps \mathbf{f}^{ref} to \mathbf{f}^{tar} and the expression for the j -th element is

$$f_j^{\text{tar}} = [\mathbf{T}(\boldsymbol{\alpha})\mathbf{f}^{\text{ref}}]_j = f^{\text{src}} \left([\boldsymbol{\nu} + \mathbf{B}\boldsymbol{\alpha}]_j \right) = \sum_{k=1}^{N_p} u_k w_k \left([\boldsymbol{\nu} + \mathbf{B}\boldsymbol{\alpha}]_j \right). \quad (2.63)$$

Define the matrix $\mathbf{W}(\boldsymbol{\alpha})$ to have entries

$$[\mathbf{W}(\boldsymbol{\alpha})]_{jk} \triangleq w_k ([\boldsymbol{\nu} + \mathbf{B}\boldsymbol{\alpha}]_j). \quad (2.64)$$

Thus, (2.58) and (2.63) can be re-expressed in the matrix-vector form

$$\begin{aligned} \mathbf{f}^{\text{ref}} &= \mathbf{W}(\mathbf{0})\mathbf{u} \\ \mathbf{T}(\boldsymbol{\alpha})\mathbf{f}^{\text{ref}} &= \mathbf{W}(\boldsymbol{\alpha})\mathbf{u}. \end{aligned} \quad (2.65)$$

2.2.3 Regularization of Image Registration

Since typical transformation estimation problem is ill-posed, a regularizer is usually used to encourage desired transformations. The choice of regularizer is based on motion priors, such as smoothness, invertibility and tissue rigidity.

Chun *et al.* [23] proposed a simple regularizer for B-spline nonrigid image registration. This penalty method yields much more realistic deformation for breathing motion than unconstrained registration methods. Moreover, it is much simpler and faster than the traditional Jacobian determinant penalty and is more memory efficient. It is computationally expensive to calculate the Jacobian determinants or its gradient due to additional B-spline interpolations of the partial derivative of a deformation. This quadratic-like regularizer enforces a sufficient condition for invertibility directly on the B-spline deformation coefficients, so it does not require additional B-spline interpolations beyond the interpolations needed for the data fitting term. It also encourages the smoothness of deformations inherently because it constrains the differences between adjacent deformation coefficients. In addition, its first and second derivatives are simple and convenient for use in optimization algorithms.

2.3 Optimization Transfer Principle

Optimization is a very important part of problems involving maximizing/minimizing cost functions, such as image reconstruction and image registration. Optimization transfer method² converts optimizing difficult cost functions to optimizing a sequence of relatively simpler surrogate func-

²This technique is also called “iterative majorization”, “majorize, minimize” and “minorize, maximize”.

tions [27, 51, 71, 72]. Mathematically, consider a minimization problem

$$\arg \min_{\mathbf{x} \in \mathcal{C}} \Psi(\mathbf{x}), \quad (2.66)$$

where $\mathcal{C} \subset \mathbb{R}^{N_p}$ is the valid parameter space (e.g., $\mathbf{x} \succeq \mathbf{0}$). At the n th iteration optimization transfer approach replaces the cost function Ψ that is difficult to minimize with a surrogate function $\phi^{(n)}(\mathbf{x})$ that is easier to minimize and satisfies the following conditions:

$$\begin{aligned} \phi^{(n)}(\mathbf{x}^{(n)}) &= \Psi(\mathbf{x}^{(n)}), \quad \mathbf{x}^{(n)} \in \mathcal{C}, \quad \text{“matched } \Psi \text{ value”} \\ \phi^{(n)}(\mathbf{x}) &\geq \Psi(\mathbf{x}), \quad \mathbf{x} \in \mathcal{C}, \quad \text{“lies above”}. \end{aligned} \quad (2.67)$$

After choosing a surrogate $\phi^{(n)}(\mathbf{x})$ in the “S-step”, a minimizer $\mathbf{x}^{(n+1)}$ of $\phi^{(n)}(\mathbf{x})$ is found in the “M-step”:

$$\mathbf{x}^{(n+1)} \triangleq \arg \min_{\mathbf{x} \in \mathcal{C}} \phi^{(n)}(\mathbf{x}). \quad (2.68)$$

Repeating the “S-step” and “M-step” iteratively, one obtains a sequence of vectors $\{\mathbf{x}^{(n)}\}$ that monotonically decrease the original cost function Ψ . The monotonicity is guaranteed by the surrogate conditions (2.67).

In general, the surrogate functions are specially designed for a cost function (or a kind of cost functions) of interest. Hunter *et al.* [51] summarized surrogate design techniques as: (a) avoiding large matrix inversions, (b) linearizing an optimization problem, (c) separating the parameters of an optimization problem, (d) dealing with equality and inequality constraints gracefully, or (e) turning a nondifferentiable problem into a smooth problem. The choice of surrogate functions is essential to the success of optimization transfer method. To achieve fast convergence rate, one wants curvatures of surrogate functions to be as small as possible while satisfying the required conditions (2.67). To achieve easy optimization in the “M-step”, one wants simple surrogate functions, such as quadratic surrogates. Since it is usually difficult to find surrogate functions that fit all the desired conditions, the choice of surrogate functions is something of an art.

CHAPTER 3

Three-Dimensional Forward and Back-Projection Methods ¹

Iterative statistical methods for 3D tomographic image reconstruction [34,74,147] offer numerous advantages such as the potential for improved image quality and reduced dose, as compared to the conventional methods such as filtered back-projection (FBP) [30]. They are based on models for measurement statistics and physics, and can easily incorporate prior information, the system geometry and the detector response.

The main disadvantage of statistical reconstruction methods is the longer computation time of iterative algorithms that are usually required to minimize certain cost functions. For most iterative reconstruction methods, each iteration requires one forward projection and one back-projection, where the forward projection is roughly a discretized evaluation of the Radon transform, and the back-projector is the adjoint of the forward projector. These operations are the primary computational bottleneck in iterative reconstruction methods, particularly in 3D image reconstruction. Forward projector methods are also useful for making digitally rendered radiographs (DRR) [14,134].

Traditional forward and back-projectors compute the intersection lengths between each tomographic ray and each image basis function. Many methods for accelerating this process have

¹This chapter is based on materials from [79–81, 84].

been proposed, *e.g.*, [56, 99–101, 108, 130, 166]. Due to the finite size of detector cells, averaging the intersection lengths over each detector cell is considered to be a more precise modeling [15, 19, 26, 126, 128, 167]. Mathematically, it is akin to computing the convolution of the footprint of each basis function and some detector blur, such as a 2D rectangular function.

Any projector method must account for the geometry of the imaging system. Cone-beam geometries are needed for axial and helical cone-beam X-ray computed tomography (CT). In 3D parallel-beam geometry projection space, there are four independent indices (u, v, φ, θ) . The ray direction is specified by (φ, θ) where φ and θ denote the azimuthal and polar angle of the ray respectively and (u, v) denote the local coordinates on a 2D area detector. In contrast, axial cone-beam projection space is characterized by three independent indices (s, t, β) and two distance parameters (D_{s0}, D_{0d}) , where β denotes the angle of the source point counter-clockwise from the y axis, (s, t) denote the detector coordinates, D_{s0} denotes the source to rotation center distance and D_{0d} denotes the isocenter to detector distance. (See Fig. 2.2). The axial cone-beam geometry is a special case of helical cone-beam geometry with zero helical pitch.

The divergence of tomographic rays in the cone-beam geometry causes depth-dependent magnification of image basis functions, *i.e.*, voxels close to the X-ray source cast larger shadows on the detector than voxels close to the detector. This complication does not appear in the parallel-beam geometry. Therefore, many existing projection and back-projection methods designed for 3D parallel-beam geometry [15, 45, 94, 126, 128] are not directly suitable for cone-beam geometry.

A variety of projection methods for 3D cone-beam geometries have been proposed [9, 14, 26, 38, 93, 120, 167]. All methods provide some compromise between computational complexity and accuracy. Among these, spherically symmetric basis functions (blobs) [93, 167] have many advantages over simple cubic voxels or other basis functions for the image representation, *e.g.*, their appearance is independent of the viewing angle. However, evaluating integrals of their footprint

functions is computationally intensive. Ziegler *et al.* [167] stored these integrals in a lookup-table. If optimized blobs are used and high accuracy is desired, the computation of forward and back-projection is still expensive due to loading a large table and the fact that blobs intersect many more tomographic rays than voxels.

Rectification techniques [120] were introduced to accelerate the computation of cone-beam forward and backward projections. Riddell *et al.* [120] resampled the original data to planes that are aligned with two of the reconstructed volume main axes, so that the original cone-beam geometry can be replaced by a simpler geometry that involves only a succession of plane magnifications. In iterative methods, resampled measurements can simplify forward and back-projection each iteration. However, resampling involves interpolation that may slightly decrease spatial resolution. Another drawback of this method is that the usual assumption of statistical independence of the original projection data samples no longer holds after rectification, since interpolation introduces statistical correlations.

The distance-driven (DD) projector [26] is a current state-of-the-art method. It maps the horizontal and vertical boundaries of the image voxels and detector cells onto a common plane such as xz or yz plane, approximating their shapes by rectangles. (This step is akin to rectification). It calculates the lengths of overlap along the x (or y) direction and along the z direction, and then multiplies them to get the area of overlap. The DD projector has the largest errors for azimuthal angles of the X-ray source that are around odd multiples of $\pi/4$, because the transaxial footprint is approximately triangular rather than rectangular at those angles.

This paper describes two new approaches for 3D forward and back-projection that we call the separable footprint (SF) projectors: the SF-TR [79] and SF-TT [80] projector. They approximate the voxel footprint functions as 2D separable functions. This approximation is reasonable for typical axial or helical cone-beam CT geometries. The separability of these footprint functions greatly

simplifies the calculation of their integrals over a detector cell and allows efficient implementation of the SF projectors. The SF-TR projector uses trapezoid functions in the transaxial direction and rectangular functions in the axial direction, whereas the SF-TT projector uses trapezoid functions in both directions. It is accurate to use rectangle approximation in the axial direction for cone-beam geometries with small cone angles ($< 2^\circ$) such as the multi-slice detector geometries, and to use trapezoid approximation for CT systems with larger cone angles ($> 10^\circ$) such as flat-panel detector geometries.

Our studies showed that both SF projector methods are more accurate than the distance-driven (DD) projector. In particular, the SF methods reduce the errors around odd multiples of $\pi/4$ seen with DD. The SF-TT projector is more accurate than the SF-TR projector for voxels associated with large cone angles. The SF-TR projector has similar computation speed with the DD projector and the SF-TT projector is about 2 times slower.

To balance computation and accuracy, one may combine the SF-TR and SF-TT projector, that is, to use the SF-TR projector for voxels associated with small cone angles such as voxels near the plane of the X-ray source where the rectangle approximation is adequate, and use the SF-TT projector for voxels associated with larger cone angles.

The organization of this paper is as follows. Section 3.1 describes the cone-beam 3D system model, and presents the analytical formula of cone-beam projections of voxel basis functions. Section 3.2 introduces the SF projectors and contrasts the SF projectors with DD projector. Section 3.3 gives simulation results, including accuracy and speed comparison between the SF-TR, SF-TT and DD projector as stand-alone modules and within iterative reconstruction. Finally, conclusions are presented in Section 3.4.

3.1 Cone-Beam Projection

3.1.1 Cone-Beam System Model

In the practice of iterative image reconstruction, rather than operating on a continuous object $f(\vec{x})$, we forward project a discretized object represented by a common basis function $\beta_0(\vec{x})$ superimposed on a $N_1 \times N_2 \times N_3$ Cartesian grid as follows:

$$f(\vec{x}) = \sum_{\vec{n}} f[\vec{n}] \beta_0\left((\vec{x} - \vec{c}[\vec{n}]) \oslash \vec{\Delta}\right), \quad (3.1)$$

where the sum is over the $N_1 \times N_2 \times N_3$ lattice that is estimated and $\vec{c}[\vec{n}] = (c_1[\vec{n}], c_2[\vec{n}], c_3[\vec{n}])$ denotes the center of the \vec{n} th basis function and $\vec{n} = (n_1, n_2, n_3) \in \mathbb{Z}^3$. The grid spacing is $\vec{\Delta} = (\Delta_1, \Delta_2, \Delta_3)$, and \oslash denotes element-wise division. We consider the case $\Delta_1 = \pm\Delta_2$ hereafter, but we allow $\Delta_1 \neq \Delta_3$, because voxels are often not cubic.

Most projection/back-projection methods use a linear model that ignores the "exponential edge gradient effect" caused by the nonlinearity of Beer's law [40, 58]. We adopt the same type of approximation here. Assume that the detector blur $h(s, t)$ is shift invariant, independent of β , and acts only along the s and t coordinates. Then the ideal noiseless projections satisfy

$$\bar{y}_\beta[s_k, t_l] = \iint h(s_k - s, t_l - t) p(s, t; \beta) ds dt, \quad (3.2)$$

where $p(s, t; \beta)$ is the 3D projection of $f(\vec{x})$ given by (2.8), and (s_k, t_l) denotes the center of detector cell specified by indices (k, l) . The methods we present are applicable to arbitrary samples (s_k, t_l) , but for simplicity of presentation and implementation we focus on the case of uniformly

spaced samples:

$$\begin{aligned}
s_k &= (k - w_s)\Delta_s, \quad k = 0, \dots, N_s - 1, \\
t_l &= (l - w_t)\Delta_T, \quad l = 0, \dots, N_t - 1, \\
w_s &= (N_s - 1)/2 + c_s, \\
w_t &= (N_t - 1)/2 + c_t,
\end{aligned} \tag{3.3}$$

where Δ_s and Δ_T denote the sample spacing in s and t respectively. The user-selectable parameters c_s and c_t denote offsets for the detector, *e.g.*, $c_s = 1/4$ corresponds to a quarter detector offset [73, 102].

Substituting the basis expansion model (3.1) for the object into (3.2) and using (2.8) leads to the linear model

$$\bar{y}_\beta[s_k, t_l] = \sum_{\vec{n}} a_{\beta}[s_k, t_l; \vec{n}] f[\vec{n}], \tag{3.4}$$

where the elements of system matrix \mathbf{A} are samples of the following cone-beam projection of a single basis function centered at $\vec{c}[\vec{n}]$:

$$a_{\beta}[s_k, t_l; \vec{n}] = F(s_k, t_l; \beta; \vec{n}), \tag{3.5}$$

where the ‘‘blurred footprint’’ function is

$$F(s_k, t_l; \beta; \vec{n}) \triangleq \iint h(s_k - s, t_l - t) q(s, t; \beta; \vec{n}) ds dt, \tag{3.6}$$

and $q(s, t; \beta; \vec{n})$ denotes the cone-beam footprint of basis function $\beta_0((\vec{x} - \vec{c}[\vec{n}]) \oslash \vec{\Delta})$, *i.e.*,

$$q(s, t; \beta; \vec{n}) = \int_{\mathcal{L}(s,t,\beta)} \beta_0((\vec{x} - \vec{c}[\vec{n}]) \oslash \vec{\Delta}) d\ell. \tag{3.7}$$

Computing the footprint of the voxel is also known as ‘‘splating’’ [153].

The goal of forward projectors is to compute (3.4) rapidly but accurately. Although the system matrix \mathbf{A} is sparse, it is impractical to precompute and store even the nonzero system matrix values

for the problem sizes of interest in cone-beam CT, so practical methods (including our proposed approach) essentially compute those values on the fly.

We focus on a simple separable model for the detector blur

$$h(s, t) = \frac{1}{r_s r_t} \operatorname{rect}\left(\frac{s}{r_s}\right) \operatorname{rect}\left(\frac{t}{r_t}\right), \quad (3.8)$$

where r_s and r_t denote the width along s and t respectively. This model accounts for the finite size of the detector elements. Note that r_s and r_t can differ from the sample spacing $s_k - s_{k-1}$ and $t_l - t_{l-1}$ to account for detector gaps.

3.1.2 Footprints of Voxel Basis Functions

We focus on cubic voxel basis functions hereafter, but one could derive analytical formulas for footprints of other basis functions. The cubic voxel basis function is given by,

$$\begin{aligned} \beta_0(\vec{x}) &= \operatorname{rect}(x) \operatorname{rect}(y) \operatorname{rect}(z) \\ &= \mathbf{1}_{\{|x| \leq 1/2\}} \mathbf{1}_{\{|y| \leq 1/2\}} \mathbf{1}_{\{|z| \leq 1/2\}}, \end{aligned} \quad (3.9)$$

where $\mathbf{1}_{\{\cdot\}}$ denotes the indicator function.

Substituting (3.9) into (3.7), the analytical formula for the cone-beam projection footprint of the \vec{n} th basis function is:

$$\begin{aligned} q(s, t; \beta; \vec{n}) &= \int_0^{L_p} \beta_0\left(\left(\vec{p}_0 + \ell \vec{e} - \vec{c}[\vec{n}]\right) \oslash \vec{\Delta}\right) d\ell \\ &= \int_0^{L_p} \mathbf{1}_{\{|d_1 + \ell e_1| \leq \Delta_1/2\}} \mathbf{1}_{\{|d_2 + \ell e_2| \leq \Delta_2/2\}} \\ &\quad \mathbf{1}_{\{|d_3 + \ell e_3| \leq \Delta_3/2\}} d\ell \\ &= a_1 \cdot a_2 \cdot a_3 \cdot [\ell_{\max} - \ell_{\min}]_+, \end{aligned} \quad (3.10)$$

where $\vec{e} = (e_1, e_2, e_3)$ was defined in (2.4), $[x]_+ \triangleq \max(x, 0)$ and

$$\begin{aligned}
\vec{d} &\triangleq \vec{p}_0 - \vec{c}[\vec{n}] = (d_1, d_2, d_3), \\
a_1 &= \begin{cases} 1_{\{|d_1| \leq \Delta_1/2\}}, & \sin\varphi = 0 \\ 1, & \sin\varphi \neq 0, \end{cases} \\
a_2 &= \begin{cases} 1_{\{|d_2| \leq \Delta_2/2\}}, & \cos\varphi = 0 \\ 1, & \cos\varphi \neq 0, \end{cases} \\
a_3 &= \begin{cases} 1_{\{|d_3| \leq \Delta_3/2\}}, & \sin\theta = 0 \\ 1, & \sin\theta \neq 0, \end{cases} \\
\ell_{\max} &= \min \{L_p, \ell_+^1, \ell_+^2, \ell_+^3\}, \\
\ell_{\min} &= \max \{0, \ell_-^1, \ell_-^2, \ell_-^3\}, \\
\ell_+^i &= \begin{cases} \max \left\{ \frac{\vec{\Delta}_i/2 - d_i}{e_i}, \frac{-\vec{\Delta}_i/2 - d_i}{e_i} \right\}, & e_i \neq 0 \\ \infty, & e_i = 0, \end{cases} \\
\ell_-^i &= \begin{cases} \min \left\{ \frac{\vec{\Delta}_i/2 - d_i}{e_i}, \frac{-\vec{\Delta}_i/2 - d_i}{e_i} \right\}, & e_i \neq 0 \\ -\infty, & e_i = 0. \end{cases}
\end{aligned} \tag{3.11}$$

For typical cone-beam geometries, polar angles θ of rays are much smaller than 90° , so there is no need to consider the case of $\cos\theta = 0$. Combining (3.5), (3.6) and (3.10) yields the “ideal” projector for cubic voxels in cone-beam CT.

3.2 Separable Footprint (SF) Projector

It would be expensive to exactly compute the true footprint (3.10) and the “blurred footprint” (3.6) for the voxel basis function on the fly, so appropriate approximations of the “blurred footprint” (3.6) are needed to simplify the double integral calculation.

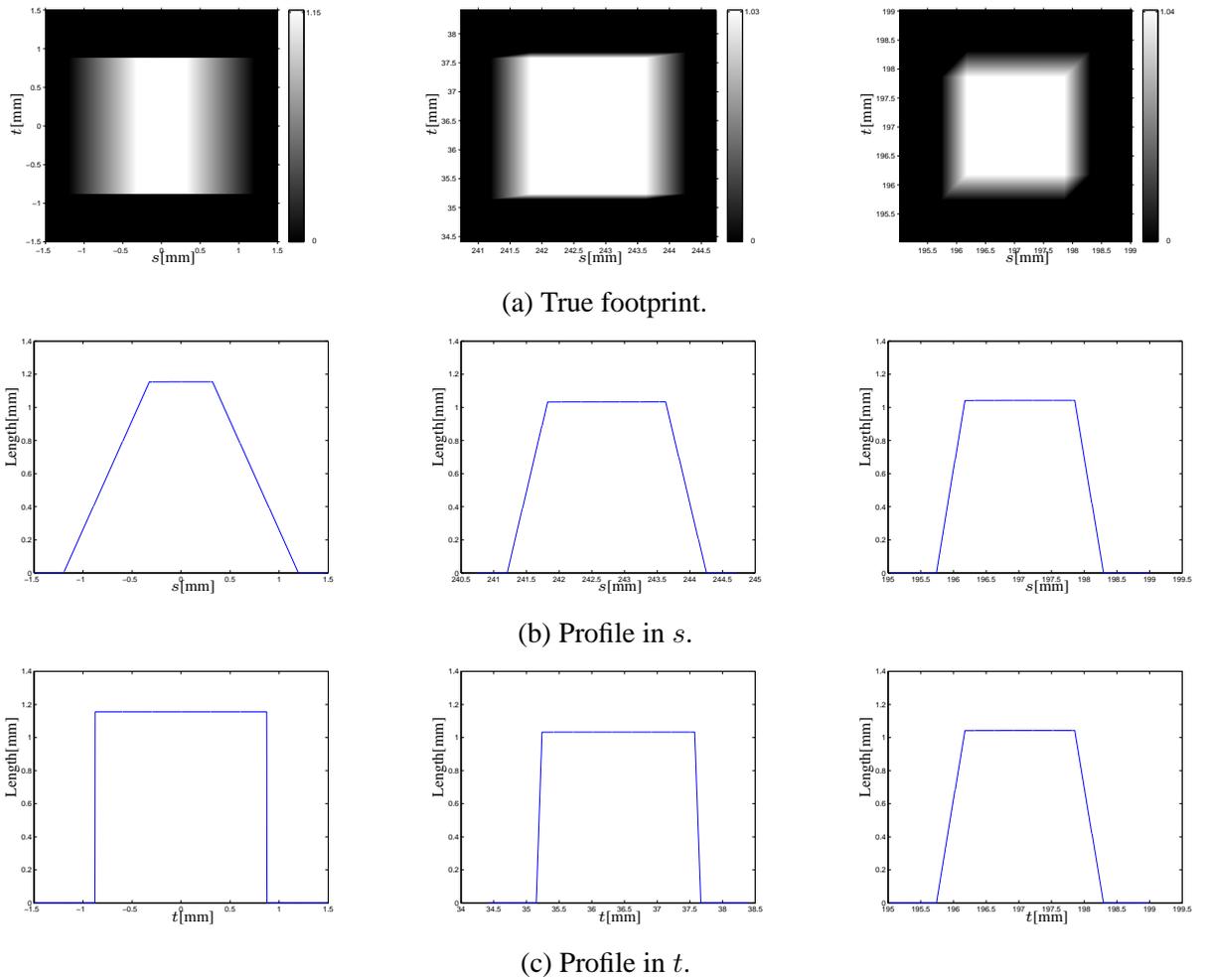


Figure 3.1: Exact footprint functions $q(s, t; \beta; \vec{n})$ and their profiles for 1mm^3 voxels centered at the origin (*left*), $(100, 150, 15)$ mm (*center*) and $(93, 93, 93)$ mm (*right*).

To explore alternatives, we simulated a flat-detector cone-beam geometry with $D_{s_0} = 541$ mm and $D_{sd} = 949$ mm. We computed cone-beam projections of voxels analytically using (3.10) at sample locations $(n\Delta_s, m\Delta_T)$ where $\Delta_s = \Delta_T = 0.001$ mm and $n, m \in \mathbb{Z}$. The left column of Fig. 3.1 shows the exact footprint function and its profiles for a voxel with $\Delta_1 = \Delta_2 = \Delta_3 = 1$ mm centered at the origin when $\beta = 30^\circ$. The center column of Fig. 3.1 shows those of a voxel centered at $(100, 150, 15)$ mm when $\beta = 0^\circ$. The azimuthal and polar angle of the ray connecting the source and this voxel center are 14.3° and 2.1° respectively. The cone angle of a typical 64-slice cone-beam CT geometry is about 2° . The right column of Fig. 3.1 shows those of a voxel centered at $(93, 93, 93)$ mm when $\beta = 0^\circ$. The azimuthal and polar angle of the ray connecting the source and this voxel center are 11.7° and 11.5° respectively. The cone angle of a typical cone-beam CT geometry with 40×40 cm² flat-panel detector is about 12° . The first two true footprints look like 2D separable functions. The third footprint is approximately separable except for small areas at the upper left and lower right corner.

Inspired by shapes of the true footprints (see Fig. 3.1), we approximate them as follows,

$$q(s, t; \beta; \vec{n}) \approx q_{ap}(s, t; \beta; \vec{n}) \triangleq l(s, t; \beta; \vec{n})q_{sf}(s, t; \beta; \vec{n}), \quad (3.12)$$

where $q_{sf}(s, t; \beta; \vec{n})$ denotes a 2D separable function with unit maximum amplitude,

$$q_{sf}(s, t; \beta; \vec{n}) \triangleq q_1(s; \beta; \vec{n})q_2(t; \beta; \vec{n}), \quad (3.13)$$

where $q_1(s; \beta; \vec{n})$ and $q_2(t; \beta; \vec{n})$ denote the approximating functions in s and t respectively. In (3.12), $l(s, t; \beta; \vec{n})$ denotes the ‘‘amplitude’’ of $q_{sf}(s, t; \beta; \vec{n})$.

For small basis functions and narrow blurs $h(s, t)$, the angles of rays within each detector cell that intersect each basis function are very similar, so $l(s, t; \beta; \vec{n})$ is much smoother than $h(s, t)$ and

$q(s, t; \beta; \vec{n})$. Substituting (3.12) into (3.6) leads to

$$\begin{aligned}
F(s, t; \beta; \vec{n}) &\approx F_{\text{sf}}(s, t; \beta; \vec{n}) \\
&\triangleq h(s, t) ** [l(s, t; \beta; \vec{n})q_{\text{sf}}(s, t; \beta; \vec{n})] \\
&\approx l(s, t; \beta; \vec{n}) [h(s, t) ** q_{\text{sf}}(s, t; \beta; \vec{n})]
\end{aligned} \tag{3.14}$$

where the inequality uses the fact that $l(s, t; \beta; \vec{n})$ is approximately a constant over each detector cell. The value $l(s_k, t_l; \beta; \vec{n})$ denotes this constant for detector cell (s_k, t_l) , and $*$ denotes 2D convolution

If the detector blur is also modeled as separable, *i.e.*,

$$h(s, t) = h_1(s)h_2(t), \tag{3.15}$$

then the blurred footprint functions (3.14) have the following separable approximation:

$$F_{\text{sf}}(s_k, t_l; \beta; \vec{n}) = l(s_k, t_l; \beta; \vec{n})F_1(s_k; \beta; \vec{n})F_2(t_l; \beta; \vec{n}), \tag{3.16}$$

where

$$\begin{aligned}
F_1(s_k; \beta; \vec{n}) &\triangleq \int h_1(s_k - s)q_1(s; \beta; \vec{n})ds \\
F_2(t_l; \beta; \vec{n}) &\triangleq \int h_2(t_l - t)q_2(t; \beta; \vec{n})dt.
\end{aligned} \tag{3.17}$$

3.2.1 Amplitude Approximation Methods

One natural choice for the amplitude function $l(\cdot)$ is the following voxel-dependent factor that we call the A3 method:

$$l(s_k, t_l; \beta; \vec{n}) = l_3(\beta; \vec{n}) \triangleq l_{\varphi_0} \cdot l_{\theta_0} \tag{3.18}$$

where

$$l_{\varphi_0} \triangleq \frac{\Delta_1}{\max\{|\cos(\varphi_0)|, |\sin(\varphi_0)|\}} \quad (3.19)$$

$$l_{\theta_0} \triangleq \frac{1}{|\cos(\theta_0)|}, \quad (3.20)$$

where $\varphi_0 = \varphi_0(\beta, \vec{n})$ and $\theta_0 = \theta_0(\beta, \vec{n})$ denote the azimuthal and polar angles of the ray connecting the source and center of the \vec{n} th voxel. They can be computed by (2.10) and (2.11). Since this voxel-dependent amplitude depends on angles (θ_0, φ_0) and β , the approximated footprint $q_{\text{ap}}(s, t; \beta; \vec{n})$ is separable with respect to s and t too. However, the dependence on voxel centers $\vec{c}[\vec{n}]$ requires expensive computation. One must compute $N_1 \times N_2 \times N_3 \times N_\beta$ different l_{θ_0} values and $N_1 \times N_2 \times N_\beta$ different l_{φ_0} values, where N_β denotes the number of projection views. In addition, computing l_{θ_0} and l_{φ_0} for each voxel at each projection view involves either trigonometric operations (\cos , \sin and \tan^{-1}) or square and square root operations to directly evaluate \cos and \sin .

To accelerate computation of the SF projector, we propose a voxel-ray-dependent amplitude named the A2 method:

$$l_2(s_k, t_l; \beta; \vec{n}) \triangleq l_{\varphi_0} \cdot l_{\theta(s_k, t_l)} \quad (3.21)$$

$$l_{\theta(s_k, t_l)} \triangleq \frac{1}{|\cos(\theta(s_k, t_l))|}, \quad (3.22)$$

where $\theta(s_k, t_l)$ given in (2.7) is the polar angle of the ray connecting the source and detector center (s_k, t_l) . There are many fewer tomographic rays ($N_s \times N_t$) than voxels in a 3D image ($N_1 \times N_2 \times N_3$) and $\theta(s_k, t_l)$ does not depend on β for flat detector geometries (see (2.7)), so using (3.21) saves substantial computation versus (3.18).

We also investigated a ray-dependent amplitude named the A1 method:

$$l_1(s_k, t_l; \beta) \triangleq l_{\varphi(s_k; \beta)} \cdot l_{\theta(s_k, t_l)} \quad (3.23)$$

$$l_{\varphi(s_k; \beta)} \triangleq \frac{\Delta_1}{\max\{|\cos(\varphi(s_k; \beta))|, |\sin(\varphi(s_k; \beta))|\}}, \quad (3.24)$$

where $\varphi(s_k; \beta)$ given in (2.6) is the azimuthal angle of the ray connecting the source and detector cell center (s_k, t_l) . For each β , there are N_s different $l_{\varphi(s_k; \beta)}$ for the A1 method and $N_1 \times N_2$ different l_{φ_0} for the A2 method.

These amplitude methods are similar to Joseph's method [56] where the triangular footprint function is scaled by $1/\max(|\cos \varphi|, |\sin \varphi|)$ for 2D fan-beam geometry. All three methods have similar accuracies, but the A3 method is much slower than the other two (see Section 3.3.1). Thus we do not recommend using the A3 amplitude in the SF projector method. Hereafter, we refer to (3.16) with either (3.21) or (3.23) as “the SF method”.

3.2.2 SF Projector with Trapezoid/Rectangle Function (SF-TR)

Inspired by the shapes of the true footprints associated with small cone angles (see the first two columns of Fig. 3.1), we approximate them as 2D separable functions with trapezoid functions in the transaxial direction and rectangular functions in the axial direction. This approximation is reasonable for typical multi-slice cone-beam geometries, where the azimuthal angles φ of rays cover the entire 360° range since the X-ray source rotates around the z axis, whereas the polar angles θ of rays are small (less than 2°) since the cone angle is small.

The approximating function in the s direction is

$$\begin{aligned}
 q_1(s; \beta; \vec{n}) &\triangleq \text{trap}(s; \tau_0, \tau_1, \tau_2, \tau_3) \\
 &= \begin{cases} \frac{s-\tau_0}{\tau_1-\tau_0}, & \tau_0 < s < \tau_1 \\ 1, & \tau_1 \leq s \leq \tau_2 \\ \frac{\tau_3-s}{\tau_3-\tau_2}, & \tau_2 < s < \tau_3 \\ 0, & \text{otherwise} \end{cases}, \quad (3.25)
 \end{aligned}$$

where τ_0, τ_1, τ_2 and τ_3 denote vertices of the trapezoid function that we choose to match the exact locations of those of the true footprint function in the s direction. They are the projected s coordinates of four corner points located at $(c_1[\vec{n}] \pm \Delta_1/2, c_2[\vec{n}] \pm \Delta_2/2)$ for all z .

The approximating function in the t direction is

$$q_2(t; \beta; \vec{n}) \triangleq \text{rect}\left(\frac{t-t_0}{w_{t0}}\right), \quad (3.26)$$

where

$$\begin{aligned}
 t_0 &\triangleq \frac{t_+ + t_-}{2}, \\
 w_{t0} &\triangleq t_+ - t_-, \quad (3.27)
 \end{aligned}$$

where t_+ and t_- denote the boundaries of the rectangular function which we choose to be the projected t coordinates of the two endpoints of the axial midline of the voxel. Those endpoints are located at $\vec{c}[\vec{n}] \pm (0, 0, \Delta_3/2)$. Given β and a point $\vec{x} = (x, y, z)$, the projected s and t coordinate of this point can be computed by (2.12) and (2.14). Since the boundaries of the separable function are determined by the projections of boundaries of the voxel basis function under the cone-beam geometry, the depth-dependent magnification is accurately modeled.

The blurred footprint functions (3.17) of this SF-TR projector are

$$F_1(s_k; \beta; \vec{n}) = \frac{1}{r_s} \gamma\left(s_k - \frac{r_s}{2}, s_k + \frac{r_s}{2}\right), \quad (3.28)$$

and

$$F_2(t_l; \beta; \vec{n}) = \frac{1}{r_t} \left[\min\left(t_l + \frac{r_t}{2}, t_+\right) - \max\left(t_l - \frac{r_t}{2}, t_-\right) \right]_+, \quad (3.29)$$

where

$$\begin{aligned} \gamma(s_1, s_2) &\triangleq \int_{s_1}^{s_2} \text{trap}(s; \tau_0, \tau_1, \tau_2, \tau_3) ds \\ &= \gamma_1(\max(s_1, \tau_0), \min(s_2, \tau_1)) \\ &\quad + \gamma_2(\max(s_1, \tau_1), \min(s_2, \tau_2)) \\ &\quad + \gamma_3(\max(s_1, \tau_2), \min(s_2, \tau_3)), \\ \gamma_1(b_1, b_2) &\triangleq \frac{1}{2(\tau_1 - \tau_0)} [(b_2 - \tau_0)^2 - (b_1 - \tau_0)^2] \mathbf{1}_{\{b_2 > b_1\}}, \\ \gamma_2(b_1, b_2) &\triangleq (b_2 - b_1) \mathbf{1}_{\{b_2 > b_1\}}, \\ \gamma_3(b_1, b_2) &\triangleq \frac{1}{2(\tau_3 - \tau_2)} [(b_1 - \tau_3)^2 - (b_2 - \tau_3)^2] \mathbf{1}_{\{b_2 > b_1\}}. \end{aligned} \quad (3.30)$$

3.2.3 SF Projector with Trapezoid/Trapezoid Function (SF-TT)

Inspired by the shape of true footprint of a voxel associated with large cone angles (see the last column of Fig. 3.1), we approximate it as a 2D separable function with trapezoid functions in *both* the transaxial and axial direction. This trapezoid approximation in axial direction is reasonable for cone-beam geometries with large cone angles ($> 10^\circ$) such as flat-panel detector geometries.

Along s , the SF-TT projector uses the same trapezoid approximation as the SF-TR projector. The trapezoid footprint and the blurred footprint are given in (3.25) and (3.28).

The approximated footprint function in t is

$$q_2(t; \beta; \vec{n}) \triangleq \text{trap}(t; \xi_0, \xi_1, \xi_2, \xi_3), \quad (3.31)$$

where ξ_0, ξ_1, ξ_2 and ξ_3 denote vertices of the trapezoid function. ξ_0 and ξ_1 are the smallest and largest one of the projected t coordinates of the lower four corners of the \vec{n} th voxel located at

$(c_1[\vec{n}] \pm \Delta_1/2, c_2[\vec{n}] \pm \Delta_2/2, c_3[\vec{n}] - \Delta_3/2)$, and ξ_2 and ξ_3 are the smallest and largest one of the projected t coordinates of the upper four corners located at $(c_1[\vec{n}] \pm \Delta_1/2, c_2[\vec{n}] \pm \Delta_2/2, c_3[\vec{n}] + \Delta_3/2)$. The blurred footprint function in t is

$$F_2(t_l; \beta; \vec{n}) = \frac{1}{r_t} \gamma \left(t_l - \frac{r_t}{2}, t_l + \frac{r_t}{2} \right), \quad (3.32)$$

where γ is given in (3.30).

By choosing the vertices of the approximating footprints to match the projections of the voxel boundaries, the approximation adapts to the relative positions of the source, voxels and detector, as true footprints do. Take a voxel centered at the origin as an example. Its axial footprint is approximately a rectangular function (see the left figure in the third row of Fig. 3.1), instead of a trapezoid function. For this voxel $\text{trap}(t; \xi_0, \xi_1, \xi_2, \xi_3)$ is almost a rectangle because $\xi_0 \approx \xi_1$ and $\xi_2 \approx \xi_3$ because ξ_0, ξ_1, ξ_2 and ξ_3 are the projected t coordinates of four axial boundaries of this voxel.

3.2.4 Implementation of SF projector

We use the system matrix model (3.5) with the separable footprint approach (3.16) for both forward and back projection, which ensures that the SF forward and back projector are exact adjoint operators of each other.

Table 3.1 summaries the SF-TR projector with the A1 amplitude method (SF-TR-A1) and with the A2 method (SF-TR-A2) for a given projection view angle β . Implementating the SF-TT projector with these two amplitude methods is similar. Implementation of the back-projector is similar, except for scaling the projections at the beginning instead of the end. The key to efficient implementation of this method is to make the inner loop over z (or equivalently over t_l) [59], because the values of $F_1(s_k; \beta; \vec{n})$ are independent of z and t_l so they are precomputed prior to

For each row $n_1 = 0, 1, \dots, N_1 - 1$ of $f[\vec{n}]$:

1. For each column $n_2 = 0, 1, \dots, N_2 - 1$:

- (a) Compute trapezoid vertices $\tau_0, \tau_1, \tau_2, \tau_3$ in (3.25) using (2.12).
- (b) Determine indices (s_k values) of detector cells intersecting with $q_1(s; \beta; \vec{n})$ in s using support $[\tau_0, \tau_3]$ and coordinates of detector cell centers and spacing given in (3.3).
- (c) Compute transaxial footprint $F_1(s_k; \beta; \vec{n})$ using (3.28) for all these s_k values and store them.
- (d) Compute l_{φ_0} using (3.19) (SF-TR-A2 only)
- (e) Compute first values of t_{\pm} in (3.27).
- (f) **For each slice $n_3 = 0, 1, \dots, N_3 - 1$:**
 - i. Determine indices (t_l values) of detector cells intersecting with $q_2(t; \beta; \vec{n})$ in t using support $[t_-, t_+]$ and coordinates of detector cell centers and spacing given in (3.3).
 - ii. **For each t_l value:**
 - A. Compute $F_2(t_l; \beta; \vec{n})$ using (3.29).
 - B. **For each s_k value:**
 - Compute projection $p(s_k, t_l; \beta; \vec{n})$ where
 - $p = f[\vec{n}]F_1(s_k; \beta; \vec{n})F_2(t_l; \beta; \vec{n})$ for SF-TR-A1,
 - $p = f[\vec{n}]l_{\varphi_0}F_1(s_k; \beta; \vec{n})F_2(t_l; \beta; \vec{n})$ for SF-TR-A2.
 - Update t_{\pm} incrementally .

Scale all the projections by $l_1(s_k, t_l; \beta)$ using (3.23) for SF-TR-A1 or by $l_{\theta(s_k, t_l)}$ using (3.22) for SF-TR-A2.

Table 3.1: Pseudo-code for the SF-TR forward projector with the A1 amplitude method (SF-TR-A1) and the A2 method (SF-TR-A2).

that loop. Because (2.14) is linear in z , the first value of t_{\pm} for a given (x, y) position can be computed prior to the inner loop over z , and subsequent values can be computed by simple incremental updates, *cf.* [22]. Thus only simple arithmetic operations and conditionals are needed for evaluating $F_2(t_l; \beta; \vec{n})$ in that inner loop; all trigonometric computations occur outside that loop. Note that this separable footprint approach does not appear to be particularly advantageous for 2D fan-beam forward and backprojection because computing the transaxial footprint $F_1(s_k; \beta; \vec{n})$ requires trigonometric operations. The compute efficiency here comes from the simple rectangular footprint approximation in the axial direction. More computation is needed for the SF-TT method because it uses trapezoids in the axial direction instead rectangles.

The implementation of amplitude $l(s_k, t_l; \beta; \vec{n})$ in (3.16) for the A1 and A2 methods are different. For the A1 method, for each β the amplitude $l_1(s_k, t_l; \beta)$ is implemented by scaling projections outside the loop over voxels since it depends on detector cells only. For the A2 method, we implemented the two terms (l_{φ_0} and $l_{\theta(s_k, t_l)}$) of $l_2(s_k, t_l; \beta; \vec{n})$ separately. We scaled the projections by $l_{\theta(s_k, t_l)}$ outside of the loop over voxels and computed l_{φ_0} outside the inner loop over z since it does not depend on z .

The SF methods require $O(N^4)$ operations for forward/back projection of a N^3 volume to/from N^3 samples of the cone-beam projections. There exist $O(N^3 \log N)$ methods for back-projection [8, 16, 17]. However, those algorithms may not capture the distance-dependent effect of detector blur incorporated in the model (3.5). In 2D one can use the Fourier Slice Theorem to develop $O(N^2 \log N)$ methods [165], but it is unclear how to generalize those to 3D axial and helical CT efficiently.

3.2.5 SF Compared with Distance-Driven (DD) Projector

To compare the SF projector with the DD projector, we formulate the DD method in the context of cone-beam system model. The original publication [26] gives only a procedural description rather than an explicit model. The DD method maps the horizontal and vertical boundaries of the image voxels and detector cells to a common plane, such as xz - or yz -plane, which we call the hv plane. It evaluates the system matrix elements (3.5) and the 3D X-ray transform of voxel basis functions (3.7) by approximations on the hv plane instead of the detector plane, *i.e.*, the st plane.

The DD projector approximates the footprint function (3.7) as follows,

$$q(s, t; \beta; \vec{n}) \approx q_{\text{dd}}(h, v; \beta; \vec{n}) = \Delta_1 \text{rect}\left(\frac{h - \tau h_0}{w_h}\right) \text{rect}\left(\frac{v - \tau v_0}{w_v}\right), \quad (3.33)$$

where

$$\begin{aligned} \tau h_0 &= \frac{\tau h_1 + \tau h_2}{2}, & w_h &= \tau h_2 - \tau h_1, & \tau h_2 &\geq \tau h_1, \\ \tau v_0 &= \frac{\tau v_1 + \tau v_2}{2}, & w_v &= \tau v_2 - \tau v_1, & \tau v_2 &\geq \tau v_1, \end{aligned}$$

where τh_1 and τh_2 denote the projected h coordinates of transaxial boundaries of a voxel centered at $\vec{c}[\vec{n}]$, and τv_1 and τv_2 denote the projected v coordinates of axial boundaries of this voxel.

Similar to the shift invariant blur function (3.8) acting along s and t axes, the DD method uses a shift variant blur function along the h and v axes that depends on β , s_k and t_l , *i.e.*,

$$h_{\text{dd}}(h, v; s_k, t_l; \beta) = l_{\text{dd}}(s_k, t_l; \beta) \frac{1}{r_h r_v} \text{rect}\left(\frac{h - dh_0}{r_h}\right) \text{rect}\left(\frac{v - dv_0}{r_v}\right), \quad (3.34)$$

where

$$\begin{aligned} l_{\text{dd}}(s_k, t_l; \beta) &= \begin{cases} \frac{1}{|\cos(\theta(s_k, t_l))| \cdot |\cos(\varphi(s_k; \beta))|}, & |\cos \beta| \geq |\sin \beta| \\ \frac{1}{|\cos(\theta(s_k, t_l))| \cdot |\sin(\varphi(s_k; \beta))|}, & |\cos \beta| < |\sin \beta| \end{cases} \\ dh_0 &= \frac{dh_1 + dh_2}{2}, & r_h &= dh_2 - dh_1, & dh_2 &\geq dh_1, \\ dv_0 &= \frac{dv_1 + dv_2}{2}, & r_v &= dv_2 - dv_1, & dv_2 &\geq dv_1, \end{aligned}$$

where dh_1 and dh_2 denote the projected h coordinates of boundaries of detector cell (s_k, t_l) in the s direction, and dv_1 and dv_2 denote the projected v coordinates of boundaries of this detector cell in the t direction.

Then the DD method approximates the ‘‘blurred footprint’’ functions (3.5) as follows,

$$\begin{aligned}
F(s_k, t_l; \beta; \vec{n}) &\approx F_{\text{dd}}(s_k, t_l; \beta; \vec{n}) \\
&= \iint h_{\text{dd}}(h, v; s_k, t_l; \beta) q_{\text{dd}}(h, v; \beta; \vec{n}) dh dv \\
&= \int_{dh_1}^{dh_2} \int_{dv_1}^{dv_2} \frac{1}{r_h r_v} l_{\text{dd}}(s_k, t_l; \beta) q_d(h, v; \beta; \vec{n}) dh dv \\
&= \frac{1}{r_h r_v} l_{\text{dd}}(s_k, t_l; \beta) [\min(\tau h_2, dh_2) - \max(\tau h_1, dh_1)]_+ \\
&\quad [\min(\tau v_2, dv_2) - \max(\tau v_1, dv_1)]_+. \tag{3.35}
\end{aligned}$$

The DD method essentially approximates the voxel footprints using rectangles in both directions on a common plane such as xz or yz plane. Due to the full 360° rotation in the transaxial direction, the rectangular function used in the DD method is a poor approximation to the true footprint function (see Fig. 3.1). It also uses the separable and shift-invariant detector blur (3.8) on the detector plane. However, the approximated separable detector blurs on the common plane based on the mapped boundaries of original detector blurs are no longer shift invariant. This appears to prevent using the inner loop over s_k that aids efficiency of the SF methods (see Table 3.1).

3.3 Simulation Results

To evaluate our proposed SF-TR and ST-TT projectors, we compared them with the DD projector, a current start-of-the-art method. We compared their accuracy and speed as single modules and within iterative reconstruction methods.

3.3.1 Forward and Back-Projector as Single Modules

We simulated an axial cone-beam flat-detector X-ray CT system with a detector size of $N_s \times N_t = 512 \times 512$ cells spaced by $\Delta_s = \Delta_T = 1$ mm with $N_\beta = 984$ angles over 360° . The source to detector distance D_{sd} is 949 mm, and the source to rotation center distance D_{s0} is 541 mm. We included a rectangular detector response (3.8) with $r_s = \Delta_s$ and $r_t = \Delta_T$.

We implemented the SF-TR and SF-TT projector in an ANSI C routine. The DD projector was provided by De Man *et al.*, also implemented as ANSI C too. All used single precision. For both the SF methods and the DD method we used POSIX threads to parallelize the operations. For the forward projector each thread works on different projection views, whereas for the back projector each thread works on different image rows (n_2).

Maximum Errors of Forward Projectors

We define the maximum error as

$$e_m(\beta; \vec{n}) = \max_{s,t \in \mathbb{R}} |F(s, t; \beta; \vec{n}) - F_{ap}(s, t; \beta; \vec{n})|, \quad (3.36)$$

where F_{ap} is any of the approximate blurred footprints by the SF-TR, SF-TT and DD methods. We generated the true blurred footprint $F(s, t; \beta; \vec{n})$ in (3.6) by linearly averaging 1000×1000 analytical line integrals of rays sampled over each detector cell. We computed the line integral of each ray by the exact method described in (3.10).

We compared the maximum errors of these forward projectors for a voxel with $\Delta_1 = \Delta_2 = \Delta_3 = 1$ mm centered at the origin. Since the voxel is centered at the origins of all axes, we choose $N_\beta = 180$ angles over only 90° rotation. Fig. 3.2 shows the errors on a logarithmic scale. We compared the proposed three amplitude methods by combining them with the SF-TR projector. The errors of the A1 method are slightly larger than those of the A2 and A3 method; the biggest

difference, at $\beta = 45^\circ$, is only 3.4×10^{-4} . The error curves of the A2 and A3 methods overlap with each other. For the SF-TT projector, we plotted only the A1 and A2 methods because the combination of the SF-TT projector and A3 method is computationally much slower but only slightly improves accuracy. For the same amplitude method, the error curves of the SF-TR and SF-TT method overlap. The reason is that the rectangular and trapezoid approximation are very similar for a voxel centered at the origin of z axis. All the SF methods have smaller errors than the DD method, *i.e.*, the maximum error of the DD projector is about 652 times larger than the proposed SF methods with the A1 amplitude, and 2.6×10^3 times larger than the SF methods with the A2 amplitude when $\beta = 45^\circ$.

Fig. 3.2 also compares the maximum errors of these forward projectors for a voxel centered at (100, 150, -100) mm. We choose $N_\beta = 720$ angles over 360° rotation. The error curves of the SF-TR projector with three amplitude methods overlap and the curves of the SF-TT projector with the A1 and A2 amplitude methods overlap with each other, demonstrating again that these three amplitude methods have similar accuracies. For voxels associated with large cone angles, the SF-TT projector is more accurate than the SF-TR projector. The maximum errors of the DD and SF-TR projector are about 13 and 3 times of that of the SF-TT projector respectively.

RMS Errors of Forward Projectors

We define the root mean square (RMS) error of the footprint approximation $F_{\text{ap}}(\cdot)$ as

$$e_r(\beta; \vec{n}) = \sqrt{\frac{1}{(s_+ - s_-)(t_+ - t_-)} \int_{s_-}^{s_+} \int_{t_-}^{t_+} (F(s, t; \beta; \vec{n}) - F_{\text{ap}}(s, t; \beta; \vec{n}))^2 ds dt}, \quad (3.37)$$

where $[s_-, s_+]$ and $[t_-, t_+]$ denote the unions of the boundaries of the true footprint $F(s, t; \beta; \vec{n})$ and approximated footprint $F_{\text{ap}}(s, t; \beta; \vec{n})$ in s and t direction.

We also compared the RMS errors of the SF-TR, SF-TT and DD projectors for these two voxels described above. Fig. 3.3 shows the errors on a logarithmic scale. The results showed similar

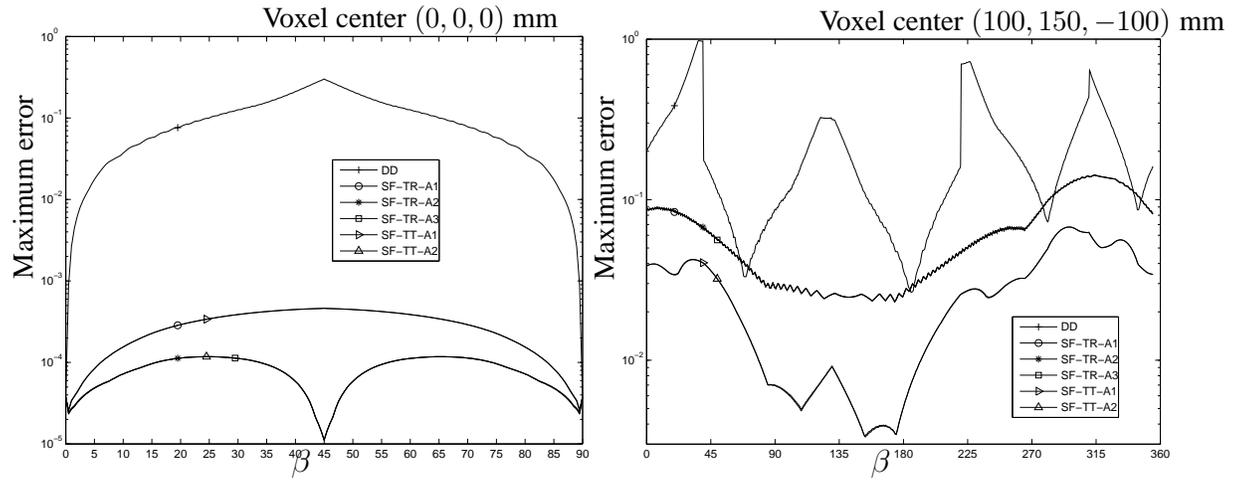


Figure 3.2: Maximum error comparison between the forward DD, SF-TR and SF-TT projector for a voxel centered at the origin (*left*) and a voxel centered at (100, 150, -100) mm (*right*).

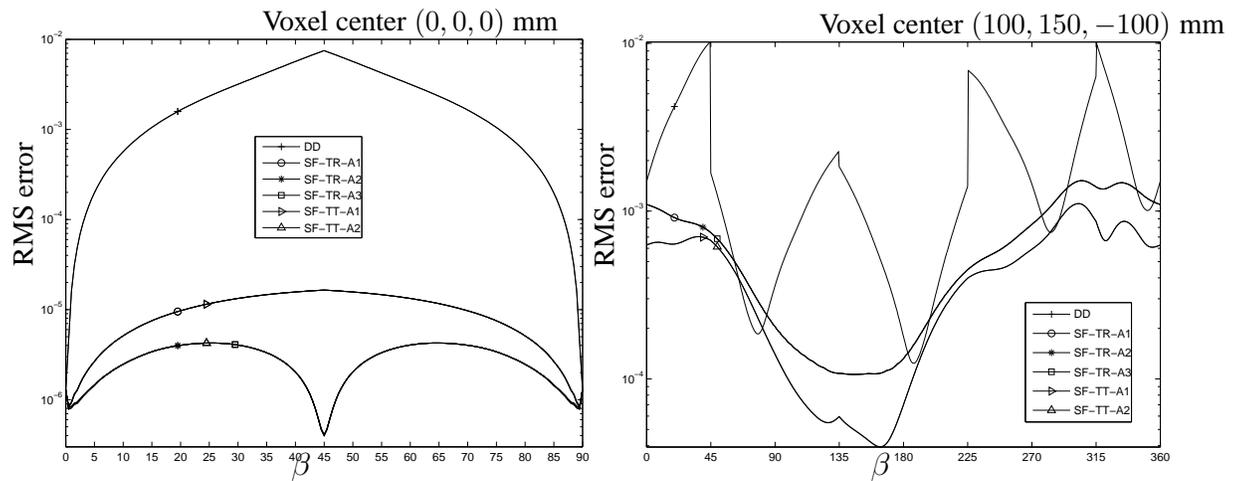


Figure 3.3: RMS error comparison between the forward DD, SF-TR and SF-TT projector for a voxel centered at the origin (*left*) and a voxel centered at (100, 150, -100) mm (*right*).

Projectors	DD	SF-TR-A1	SF-TR-A2	SF-TR-A3	SF-TT-A1	SF-TT-A2
Forward time (seconds)	46	35	35	59	91	91
Backward time (seconds)	49	44	45	63	92	93

Table 3.2: Speed comparison of DD, SF-TR and SF-TT forward and back projectors.

trends as the maximum errors. The three amplitude methods have similar accuracies. For voxels associated with small cone angles, the SF-TT and SF-TR projector have similar accuracies, while for voxels associated with large cone angles, the SF-TT projector is more accurate than the SF-TR projector. All the SF methods have smaller errors than the DD method. For the voxel centered at the origin, the RMS error of the DD projector is about 456 times larger than the proposed SF methods with the A1 amplitude, and 1.7×10^3 times larger than the SF methods with the A2 amplitude when $\beta = 45^\circ$. For the other voxel, the RMS errors of the DD and SF-TR projector are about 8 and 5 times of that of the SF-TT projector respectively.

Speed of Forward and Back-Projectors

We compared computation times of the DD, SF-TR and SF-TT forward and backward projectors using an image with a size of $N_1 = 512, N_2 = 512, N_3 = 128$ and a spacing of $\Delta_1 = \Delta_2 = \Delta_3 = 0.5$ mm in the x, y, z direction respectively. We evaluated the elapsed time using the average of 5 projector runs on a 8-core Sun Fire X2270 server with 2.66 GHz Xeon X5500 processors. Because of the “hyperthreading” of these Nehalem cores, we used 16 POSIX threads. (We found that using 16 threads reduced computation time by only about 10% compared to using 8 threads.)

Table 3.2 summarizes the computation times. For the SF-TR projector, the A1 and A2 amplitude methods have similar speed, but the A3 method is about 50% slower. The computation times of the SF-TR and DD projector are about the same, whereas the SF-TT projector is about 2 times slower. Although execution times depend on code implementation, we expect SF-TR and DD to have fairly similar compute times because the inner loop over z involves similar simple arithmetic

operations for both methods.

3.3.2 Forward and Back-Projector within Iterative Reconstruction

Since the A1 and A2 amplitude methods have similar accuracy and speed, we compared the DD and SF projectors (SF-TR and SF-TT) within iterative image reconstructions with the A1 method for simplicity.

SF-TR vs. DD

In many cases, the region of interest (ROI) needed for diagnosis is much smaller than the scanner field of view (FOV). ROI reconstruction can save computation time and memory. Ziegler *et al.* [168] proposed the following approach for iterative reconstruction of a ROI.

1. Iterative reconstruction of the whole FOV, yielding an initial estimate $\hat{\mathbf{x}}_{\text{FOV}}$ of \mathbf{x}_{FOV} which is the vector of basis coefficients of the object $f(\vec{x})$, *i.e.*, $f[\vec{n}]$ in (3.1).
2. Define $\hat{\mathbf{x}}_{\text{FOV}}^m = \hat{\mathbf{x}}_{\text{FOV}} \cdot \mathbf{m}$ where $\mathbf{m} = (m_1, \dots, m_p)$ with $0 \leq m_j \leq 1$ ($j = 1, \dots, p$) is a mask vector setting the estimated object, inside the ROI to zero and providing a smooth transition from the ROI to the remaining voxels.
3. Compute $\mathbf{p}_{\text{out}} = \mathbf{A}\hat{\mathbf{x}}_{\text{FOV}}^m$ which is the forward projection of the masked object $\hat{\mathbf{x}}_{\text{FOV}}^m$.
4. Compute the projection of ROI, $\mathbf{p}_{\text{roi}} = \mathbf{y} - \mathbf{p}_{\text{out}}$ where \mathbf{y} is the measured data.
5. Iterative reconstruction of the ROI only from \mathbf{p}_{roi} . Due to the transition zone, the region of this reconstruction needs to be extended slightly from the predetermined ROI.

This method requires accurate forward and back projectors. Errors in step 2, where re-projection of the masked image is computed, can greatly affect the results of subsequent iterative ROI reconstruction. Moreover, for general iterative image reconstruction, even small approximation errors

might accumulate after many iterations. We evaluated the accuracy of our proposed SF-TR projector and the DD projector in this iterative ROI reconstruction method.

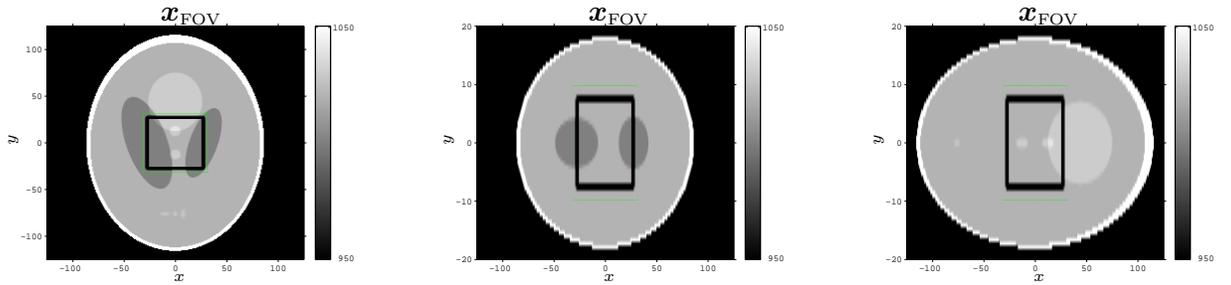
We simulated the geometry of a GE LightSpeed X-ray CT system with an arc detector of 888 detector channels for 64 slices ($N_s = 888, N_t = 64$) by $N_\beta = 984$ views over 360° . The size of each detector cell was $\Delta_s \times \Delta_T = 1.0239 \times 1.0964\text{mm}^2$. The source to detector distance was $D_{sd} = 949.075\text{mm}$, and the source to rotation center distance was $D_{s0} = 541\text{mm}$. We included a quarter detector offset in the s direction to reduce aliasing.

We used a modified 3D Shepp-Logan digital phantom that has ellipsoids centered at the $z = 0$ plane to evaluate the projectors. The brain-size field of view (FOV) was $250 \times 250 \times 40\text{mm}^3$, sampled into $256 \times 256 \times 64$ voxels with a coarse resolution of $0.9766 \times 0.9766 \times 0.6250\text{mm}^3$.

We simulated noiseless cone-beam projection measurements from the Shepp-Logan phantom by linearly averaging 8×8 analytical rays [60, p. 104] sampled across each detector cell. Noiseless data is used because we want to focus on projector accuracy. We scaled the line integrals by a chosen factor to set their maximum value to about 5.

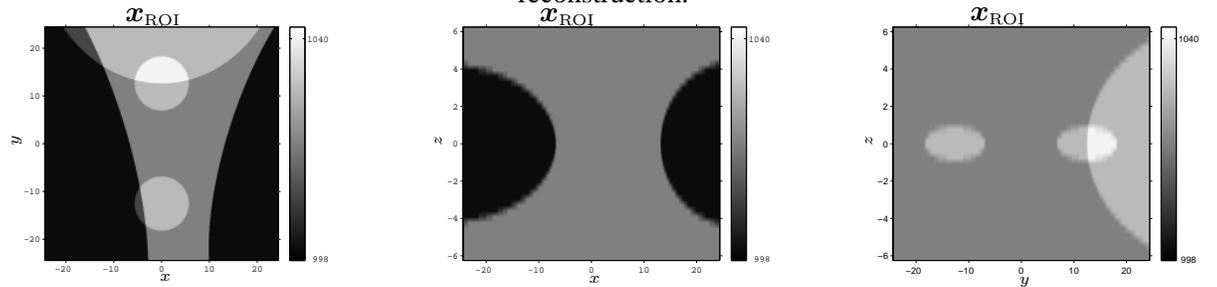
We chose a ROI centered at the rotation center that covered about $48.8 \times 48.8 \times 12.5\text{mm}^3$ ($50 \times 50 \times 20$ voxels with the coarse resolution). The transition zone surrounds the ROI, and covers about $13.7 \times 13.7 \times 5\text{mm}^3$ ($14 \times 14 \times 8$ voxels with the coarse resolution). To construct masked images $\hat{\mathbf{x}}_{\text{FOV}}^m$, we removed the ROI and smoothly weighted the voxels corresponding to the transition zone by a 3D separable Gaussian function. Fig. 3.4 shows different views of \mathbf{x}_{FOV} with the transition zone superimposed on it in the first row.

We implemented iterative image reconstruction of the entire FOV with these two projector/ back-projector methods. We ran 300 iterations of the conjugate gradient algorithm, initialized with reconstruction by the FDK method [30], for the following penalized weighted least-squares cost



(a) FOV images.

The black rectangular box shows the transition zone. The green lines show the region of ROI reconstruction.



(b) ROI images.

Figure 3.4: Shepp-Logan digital phantoms in Hounsfield units. The first, second and third columns show axial, coronal and sagittal views respectively.



Figure 3.5: Axial views of FOV images $\hat{x}_{\text{FOV}}^{\text{SF-TR}}$ and $\hat{x}_{\text{FOV}}^{\text{DD}}$ reconstructed by the iterative method (PWLS-CG) using the SF and DD method respectively. *Left*: SF-TR projector; *Right*: DD projector.

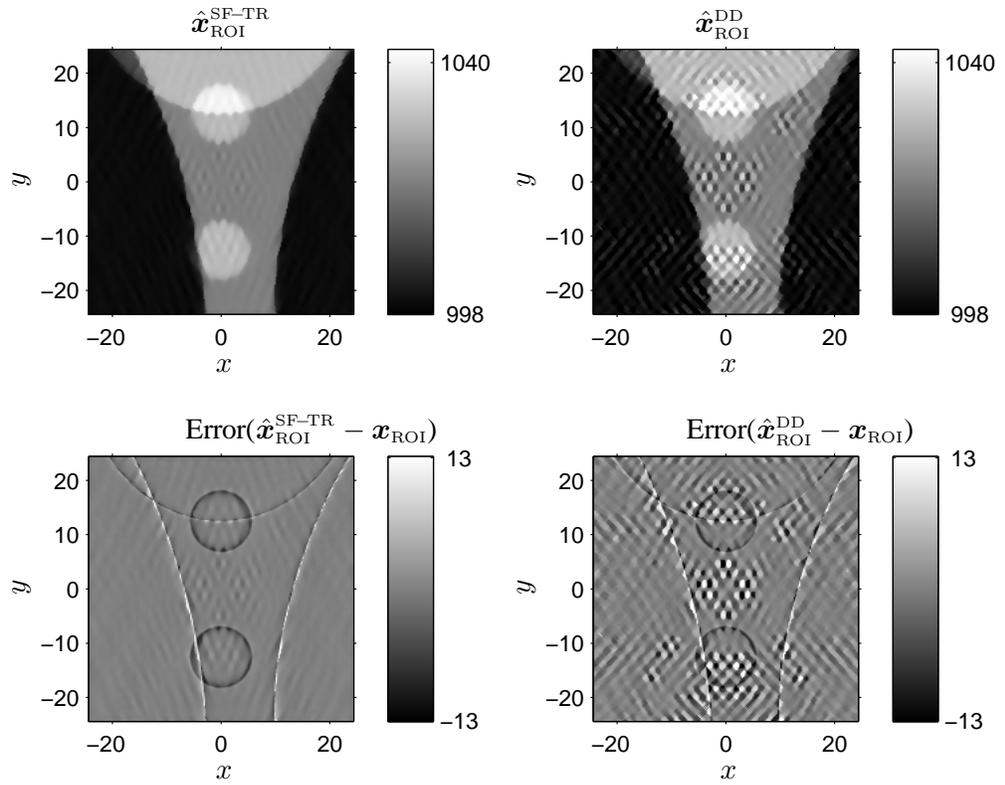


Figure 3.6: Axial views of ROI images $\hat{\mathbf{x}}_{\text{ROI}}^{\text{SF-TR}}$ and $\hat{\mathbf{x}}_{\text{ROI}}^{\text{DD}}$ reconstructed by the iterative method (PWLS-CG) using the SF-TR and DD method respectively. *Left:* SF-TR projector; *Right:* DD projector.

function with an edge-preserving penalty function (PWLS-CG):

$$\Phi(\mathbf{x}_{\text{FOV}}) = \sum_i w_i \frac{1}{2} (y_i - [\mathbf{A}\mathbf{x}_{\text{FOV}}]_i)^2 + \beta R(\mathbf{x}_{\text{FOV}}) \quad (3.38)$$

$$R(\mathbf{x}_{\text{FOV}}) = \sum_k \psi([\mathbf{C}\mathbf{x}_{\text{FOV}}]_k), \quad (3.39)$$

where y_i is the negative log of the measured cone-beam projection, w_i values are statistical weighting factors, \mathbf{A} is the system matrix, \mathbf{C} is a differencing matrix and $\psi(t)$ is the potential function.

We used the hyperbola:

$$\psi(t) = \frac{\delta^2}{3} \left(\sqrt{1 + 3 \left(\frac{t}{\delta} \right)^2} - 1 \right). \quad (3.40)$$

For this simulation, we used $w_i = \exp(-y_i)$, $\beta = 4$ and $\delta = 5$ Hounsfield units (HU).

Fig. 3.5 shows axial views of the reconstructed images $\hat{\mathbf{x}}_{\text{FOV}}^{\text{SF-TR}}$ and $\hat{\mathbf{x}}_{\text{FOV}}^{\text{DD}}$ by the iterative method (PWLS-CG) using the SF-TR and DD method respectively. We computed the maximum error, $\max_j |\hat{x}_j - x_j|$, and root-mean-square (RMS) error, $\sqrt{\frac{1}{N} \sum_{j=1}^N (\hat{x}_j - x_j)^2}$. The maximum and RMS errors of $\hat{\mathbf{x}}_{\text{FOV}}^{\text{SF-TR}}$ and $\hat{\mathbf{x}}_{\text{FOV}}^{\text{DD}}$ are close because the errors are dominated by the axial cone-beam artifacts due to the poor sampling (not truncation) at the off-axis slices, but the DD method causes artifacts that are obvious around the top and bottom areas. Similar artifacts of the DD method were reported in [145].

We applied the PWLS-CG iterative method mentioned above with $\beta = 1$ and $\delta = 1\text{HU}$ to reconstruct estimated ROI images $\hat{\mathbf{x}}_{\text{ROI}}^{\text{SF-TR}}$ and $\hat{\mathbf{x}}_{\text{ROI}}^{\text{DD}}$ of $256 \times 256 \times 64$ voxels with a fine resolution of $0.2441 \times 0.2441 \times 0.3125\text{mm}^3$. The domains of $\hat{\mathbf{x}}_{\text{ROI}}^{\text{SF-TR}}$ and $\hat{\mathbf{x}}_{\text{ROI}}^{\text{DD}}$ covered the ROI and transition zone (see Fig. 3.4). For this image geometry, we also generated a Shepp-Logan reference image \mathbf{x}_{ROI} from the same ellipsoid parameters used to generate \mathbf{x}_{FOV} . Fig. 3.4 shows different views of \mathbf{x}_{ROI} in the second row. The fine sampling of \mathbf{x}_{ROI} is 1/4 and 1/2 of the coarse sampling of \mathbf{x}_{FOV} in the transaxial and axial direction respectively, and has a size of $200 \times 200 \times 40$.

Fig. 3.6 shows the axial view of reconstructed images $\hat{\mathbf{x}}_{\text{ROI}}^{\text{SF-TR}}$ and $\hat{\mathbf{x}}_{\text{ROI}}^{\text{DD}}$ by the iterative method (PWLS-CG) using the SF-TR and DD projector. The maximum errors are 20 HU and 105 HU for the SF and DD method respectively and the RMS errors are 1.6 HU and 2.8 HU. The SF-TR projector provides lower artifact levels than the DD projector. The rectangle approximation in the transaxial direction of the DD method resulted in larger errors in the reprojection step and caused more errors when resolution changed from coarse to fine. The rectangle approximation basically blurs corners of image voxels, and the level of blur varies for different image voxel sizes.

We also reconstructed full FOV images (not shown) at a fine resolution, *i.e.*, $1024 \times 1024 \times 128$ voxels with a spacing of $0.2441 \times 0.2442 \times 0.3125\text{mm}^3$. There were no apparent artifacts in both reconstructed images using the SF-TR and DD method and the maximum and RMS errors were similar. It seems that the aliasing artifacts in the reconstruction by the DD method were removed by fine sampling [161, 162]. For smaller transaxial voxel sizes, the difference between the rectangular (DD method) and trapezoid (SF-TR) approximation becomes less visible.

SF-TR vs. SF-TT

We compared the SF-TR and SF-TT projectors by reconstructing an image under an axial cone-beam CT system with largest cone angle of 15° or so using these two methods [80]. We expected to see differences in some off-axis slices of the reconstructed images because the trapezoid approximation of the SF-TT method is more realistic than the rectangle approximation of the SF-TR method especially for voxels far away from the origin. Nevertheless, we did not see obvious visual difference, and the maximum and RMS errors were similar. It appears that the axial cone-beam artifacts due to poor sampling (not truncation) at the off-axis slices dominate other effects in the reconstructed images, such as the errors caused by rectangle approximation. Further research will evaluate these two projectors within iterative reconstruction methods under other CT geometries

where the off-axis sampling is better, such as helical scans, yet where the cone angle is large enough to differentiate the SF-TR and SF-TT method .

3.4 Conclusion

We presented two new 3D forward and back projector for X-ray CT: SF-TR and SF-TT. Simulation results have shown that the SF-TR projector is more accurate with similar computation speed than the DD projector, and the SF-TT projector is more accurate but computationally slower than the SF-TR projector. The DD projector is particularly favorable relative to other previously published projectors in terms of the balance between speed and accuracy. The SF-TR method uses trapezoid functions in the transaxial direction and rectangular functions in the axial direction, while the SF-TT method uses trapezoid functions in both directions. The rectangular approximation in the axial direction is adequate for CT systems with small cone angles, such as the multi-slice geometries. The trapezoid approximation is more realistic for geometries with large cone angles, such as the flat-panel detector geometries. To balance accuracy and computation, we recommend to combine the SF-TR and SF-TT method, which is to use the SF-TR projector for voxels corresponding to small cone angles and to use the SF-TT projector for voxels corresponding to larger cone angles.

The model and simulations here considered an ideal point source. For a finite sized X-ray source there would be more blur and it is possible that the differences between the SF and DD methods would be smaller.

Approximating the footprint functions as 2D separable functions is the key contribution of this approach. Since the separability greatly simplifies the calculation of integrals of the footprint functions, using more accurate functions in the transaxial and axial direction is possible without complicating significantly the calculations.

The computational efficiency of the SF methods rely on the assumption that the vertical (t) axis of the detector plane is parallel to the rotation axis. If the detector plane is slightly rotated then slight interpolation would be needed to resample onto coordinates that are parallel to the rotation axis.

Although we focused on voxel basis functions in this paper, the idea of 2D separable footprint approximation could also be applied to other basis functions with separability in the axial and transaxial directions, with appropriate choices of functions.

Further research will address the implementation of the SF projectors based on graphics processing unit (GPU) programming techniques [110, 134, 154] to improve the speed. Implementing the SF projectors on field-programmable gate array (FPGA) is another future research topic [64, 65].

CHAPTER 4

Two-Material Decomposition from A Single CT Scan Using Statistical Image Reconstruction ¹

Clinical CT scans are taken at the diagnostic range of X-ray energies of 30 to 150 Kev [87], and images of linear attenuation coefficients at these energies are reconstructed. However, attenuation coefficients at higher energies are often required in medical applications. Radiotherapy desires attenuation images in the Mev range for precise dose calculation and segmentation of anatomy for treatment planing [10, 127]. PET/CT scanners demand CT transmission images at 511 Kev for attenuation correction of PET emission images [66, 104]. Thus the attenuation coefficients at a lower energy must be converted to a higher energy. There is no one-to-one correspondence between attenuation coefficients at two different energies because attenuation coefficients depend on materials' densities and atomic numbers in a mixture [66, 127]. Two materials with different atomic numbers may have similar attenuation coefficients at one energy, but different coefficients at another energy. Basis material decomposition has been proposed to solve this conversion problem. An accurate image of attenuation coefficients of a mixture at any effective energy of interest can be synthesized by combining component images separated at other energies.

¹This chapter is based on material from [85].

Dual-energy (DE) CT methods, pioneered by Alvarez and Macovski *et al.* [5, 6, 88, 92, 139], are the most predominant approaches for reconstructing two basis materials (*e.g.*, soft-tissue and bone). They decomposed the energy dependence of attenuation coefficients into two components, one approximated the photoelectric interaction and another approximated Compton scattering, and separated these two components from two sets of measurements at two different source energies. Using singular value decomposition, Lehmann and Alvarez [75] showed that two basis materials are sufficient to present the energy dependence information in a mixture if operating far from the K-edges of any component material. Although DECT methods were originally proposed in the late 1970s and early 1980s, only recently DECT scanners became clinically available with technological developments, such as fast kVp-switching, dual-source CT and dual-layer detectors. These new techniques have brought renewed interest in DECT [39, 48, 50, 61, 66, 78, 86, 103, 104, 112, 112, 144, 160].

FBP based methods dominated DECT reconstruction until the early 1990s, when a few algebraic iterative methods [68, 91, 97] were proposed. FBP methods are known for amplifying noise due to the use of non-ideal ramp filter. Algebraic methods can improve the accuracy relative to FBP methods, but they do not account for noise statistics. In contrast, statistical image reconstruction methods [35, 37, 48, 104, 144] based on the physical model of the CT system and a statistical model can obtain lower noise images.

However, typical DECT methods require either two scans or specialized scanner designs [146]. Several methods have been developed to reconstruct two basis materials from one CT scan of a single tube voltage setting. Ritchings and Pullan [121] have described a technique for acquiring dual energy data by filtering alternate detectors. Although this is a single scan method, it is dose inefficient, since the photons stopped by the filters contribute to patient dose but not to the signal, and the angular spatial resolution is decreased by a factor of two. More recently, Taschereau *et*

al. [146] retrofitted a preclinical microCT scanner through a filter wheel to alternate two beam filters between successive projections. One filter provides a low energy beam while the other filter provides a high energy beam. MicroCT scanners rotate relatively slower than typical CT scanners, so alternating beam filters may not be feasible for typical CT scanners. Rutt and Fenster [123] proposed a split filter technique which requires only a simple addition of pre-patient filters to a conventional scanner. A split filter has two parts, a right half-filter and a left half-filter, and these two parts consist of different materials or thicknesses. Split filter technique screens the two halves of the X-ray beam differently at one projection angle. Using 360° rotation of the X-ray source, the whole scanned object is exposed to two incident spectra with different effective energies. For the 2D fan-beam geometry, every path in the patient cross section can easily be measured with two different beam filtration during a single scan. Rutt and Fenster first separated the data obtained through the right half-filter from those through the left half-filter to form two sets of measurements at dual energies, and then followed the decomposition method proposed by Alvarez and Macovski *et al.* in [5]. They presented a general derivation of noise analysis in dual energy analysis and applied it to determine the optimal parameters for the split-filter scans. To simplify the split-filter design optimization, they assumed the right half-fan was unfiltered in all calculations. They concluded that 1 mm copper across half the fan beam with no filtration of the other half is the optimal filtration parameters for photoelectric/Compton imaging.

We propose a statistical penalized weighted least-squares (PWLS) method with edge-preserving regularization to reconstruct two basis materials (*e.g.*, soft tissue and bone) from a single-energy CT scan. This method exploits the incident spectra difference of rays created by filtration, such as split and bow-tie filters.

Separation of measurements obtained by the right half-filter and left half-filter is not required for the statistical method since it models the physics of CT systems. In addition, for axial and heli-

cal cone beam CT geometries, most paths in the patient are not measured by exactly two different incident spectra created by split filters. The choice of material and thickness for the right and left half-filter should provide sufficient difference between the right and left spectra to minimize the ill-conditioning of this material decomposition problem. Taschereau *et al.* [146] alternated a filter of 2 mm aluminum and the other filter of 200 μm molybdenum between successive projections to obtain dual-energy measurements. The Al filter screened out very low energy X-rays, while the Mo filter greatly reduced the spectrum above 20 KeV because its K absorption edge is at 20 KeV. The Al filter produced “soft” beams while the Mo filter produced “hard” beams. They proved that the spectra between the “soft” and “hard” beams have sufficient difference for the dual-energy material decomposition problem. We borrowed this design for our split filter. For a simulated spectrum with X-ray tube voltage of 80 kVp, this split filter produced “soft” beams with effective energies of 49 keV and “hard” beams with effective energies of 58 keV.

Bow-tie filters are commonly employed in CT scanners to reduce intensity variations across detector elements. They harden the portions of the X-ray beam that will pass through the thinner parts of the object by filtering out the lower-energy components. Due to different extent of filtration to each radiation ray in the beam and the rotation of the X-ray source, the whole scanned object is exposed to incident spectra with different effective energies. Most image reconstruction methods, including the DECT methods, assume the incident spectrum is a constant for all radiation rays in the presence of bow-tie filters. In contrast, we use these differences to reconstruct two basis materials from CT measurements at a single energy. We simulated a typical body bow-tie filter that consists of aluminum, graphite, copper and titanium. For a simulated spectrum with X-ray tube voltage of 80 kVp, the effective energies are 49 keV and 56 keV for the rays at the edge and center of the bow-tie filter respectively.

We proposed an optimization transfer method with separable quadratic surrogates to monoton-

ically decrease the PWLS cost function. We first reconstructed the bone-corrected FBP images using the Joseph and Spital method [57], and separated the soft tissue and bone components by a threshold to initialize the iterative optimization algorithm. We used the ordered subsets approach to accelerate the convergence to a good local minimum.

Experiments showed that the proposed PWLS method was able to reconstruct soft tissue and bone components from a single-energy scan acquired with split and bow-tie X-ray filters. The split and bow-tie filtration had similar results. For both filtration methods, the RMS errors of the soft tissue and bone images reconstructed by the PWLS method were about 75% and 60% of those of the traditional non-iterative JS- FBP method respectively. For the density (sum of soft tissue and bone) and linear attenuation coefficient images at 511 keV, the RMS errors of the PWLS with both filtration were 40% of those by the JS-FBP method. The PWLS method reduced the beam-hardening artifacts much more effectively than the JS-FBP method which is known for mitigating these artifacts. The PWLS method also produced lower noise.

The organization of this chapter is as follows. Section 4.1 describes the differential filtration technique for producing spectra with different effective energies. Section 4.2 introduces the proposed PWLS reconstruction method. Section 4.3 discusses the optimization method with separable quadratic surrogates. Section 4.4 shows simulation results using split and bow-tie filters.

4.1 Differential Filtration

X-ray filters attenuate X-rays by different amounts depending on the X-ray photon energy. Given a X-ray spectrum, filters of different materials or thicknesses create different exiting spectra. We investigate spectra differences produced by split and bow-tie filters.

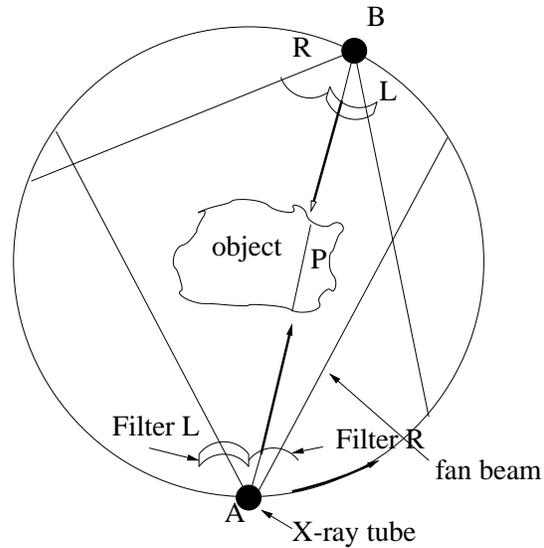


Figure 4.1: Schematic diagram of a fan-beam CT scanner with two positions of X-ray source and split-filter assembly.

4.1.1 Split Filtration

The split filter technique was proposed by Rutt and Fenster [123] in 1980. A split-filter is composed of two parts, a right half-filter R and a left half-filter L. Filters R and L are made of different materials or thicknesses. Split-filters are located near the X-ray tube to filter two halves of the X-ray beam differently. Fig. 4.1 shows a schematic diagram of a fan beam CT scanner with two positions of X-ray tube and split-filter assembly. The path P in the object is passed through by two rays. One is the ray screened by filter R at position A, the other is the ray screened by filter L at position B from the opposite side of the object. Similarly, each path in the cross section will be filtered differently since two opposing rays originated from opposite halves of the fan pass through it. Most DECT methods acquire dual energy information by two scans or specialized scanner designs, such as fast kVp-switching, dual-source CT or dual-layer detectors. In contrast, the split-filter technique produces two spectra with different effective energies by two different beam filters. The split-filter technique is independent of the CT scanner geometry, and therefore can be applied

to any existing scanner that performs 360° scans. The implementation of this technique requires only the attachment and alignment of the filters between the X-ray tube and the patient.

The choice of material and thickness for the right and left filter should provide sufficient differences between the right and left spectra to produce dual-energy measurements. We use 2 mm aluminum across half the fan beam with 200 μm molybdenum of the other half. The Al filter screens out very low energy X-rays, while the Mo filter greatly reduces the spectrum above 20 KeV because its K absorption edge is at 20 KeV [146]. The Al filter produces “soft” beams while the Mo filter produces “hard” beams. Fig. 4.2 shows the spectra created by this design of split filter for a simulated spectrum with X-ray tube voltage of 80 kVp. The effective energies are 49 keV and 58 keV for the “soft” and “hard” spectrum respectively.

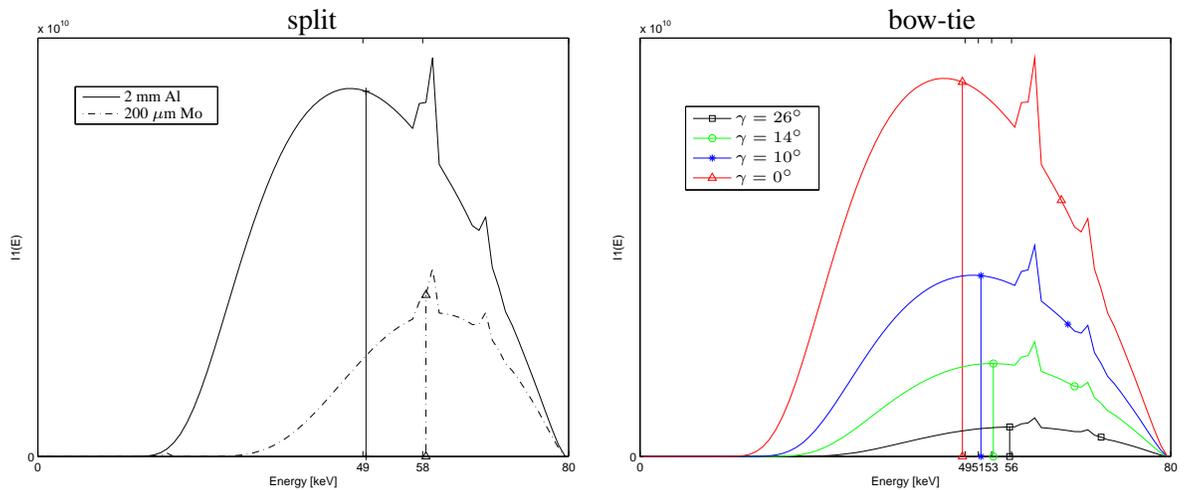


Figure 4.2: *Left:* Spectra produced by a split filter for a simulated spectrum with X-ray tube voltage of 80 kVp. The split filter consists of 2 mm aluminum producing “soft” spectra and 200 μm molybdenum producing “hard” spectra. The effective energies are 49 keV and 58 keV respectively. *Right:* Sample spectra at four fan angles screened by a body bow-tie filter. γ denotes the fan angle. The effective energies are 49, 51, 53 and 56 keV for the spectra at 26° , 14° , 10° and 0° respectively.

4.1.2 Bow-tie Filtration

Bow-tie filters are commonly used in CT scanners to minimize intensity variations across detector cells. They harden the portions of the X-ray beam that will pass through the thinner parts of the object by filtering out the lower-energy components. The thicknesses of bow-tie filters increase gradually from the center to the edge. The X-ray spectra coming out from different locations along bow-tie filters vary due to different extents of filtration. Using 360° rotation of the source, the scanned object is exposed to incident spectra with different effective energies. We simulated a typical body bow-tie filter that consists of aluminum, graphite, copper and titanium. Fig. 4.2 also shows spectra at four fan angles (26°, 14°, 10° and 0°) filtered by the simulated bow-tie filter. Their effective energies are 49, 51, 53 and 56 keV respectively.

4.2 Penalized Weighted Least-Squares (PWLS) Reconstruction

Clinthorne and Sukovic [24, 141–144] have investigated PWLS approaches for dual-energy and triple-energy CT reconstruction. They assumed monoenergetic source spectra. Huh and Fessler [48, 50] proposed a PWLS method for DECT that included a complete polyenergetic source spectrum model. We adopt this polyenergetic PWLS method to estimate two basis materials from a single scan, *i.e.*, $M_0 = 1$ and $L_0 = 2$ in (4.1).

Combining the general measurement model (2.1) and the object model (2.24), the mean of measurements can be represented as follows,

$$\bar{y}_{im} = I_{im} e^{-f_{im}(\mathbf{s}_i(\mathbf{x}))} + r_{im}, \quad (4.1)$$

for $m = 1, \dots, M_0$ and $l = 1, \dots, L_0$, where

$$f_{im}(\mathbf{s}_i) \triangleq -\log v_{im}(\mathbf{s}_i) \quad (4.2)$$

$$v_{im}(\mathbf{s}_i) \triangleq \int p_{im}(\mathcal{E}) e^{-\beta(\mathcal{E}) \cdot \mathbf{s}_i} d\mathcal{E} \quad (4.3)$$

$$p_{im}(\mathcal{E}) \triangleq I_{im}(\mathcal{E}) / I_{im} \quad (4.4)$$

$$I_{im} \triangleq \int I_{im}(\mathcal{E}) d\mathcal{E}. \quad (4.5)$$

The nonlinear function $f_{im}(\mathbf{x})$ characterizes the beam hardening caused by polychromatic source spectra. The total intensity I_{im} for the m th energy scan and the i th ray is defined in (4.5). The sinogram vector $\mathbf{s}_i(\mathbf{x})$ and mass attenuation vector $\beta(\mathcal{E})$ are defined as

$$\begin{aligned} \mathbf{s}_i(\mathbf{x}) &\triangleq (s_{i1}(\mathbf{x}), \dots, s_{iL_0}(\mathbf{x})) \\ \beta(\mathcal{E}) &\triangleq (\beta_1(\mathcal{E}), \dots, \beta_{L_0}(\mathcal{E})) \\ s_{il}(\mathbf{x}) &\triangleq [\mathbf{A}\mathbf{x}_l]_i, \end{aligned} \quad (4.6)$$

where \mathbf{A} denotes the $N_d \times N_p$ system matrix with entries

$$a_{ij} \triangleq \int_{\mathcal{L}_i} b_j(\vec{x}) d\ell. \quad (4.7)$$

The image vector is $\mathbf{x} = (\mathbf{x}_1, \dots, \mathbf{x}_l, \dots, \mathbf{x}_{L_0}) \in \mathbb{R}^{N_p \times L_0}$ for $\mathbf{x}_l = (x_{l1}, \dots, x_{lj}, \dots, x_{lN_p})$. Given noisy measurements Y_{im} , the f_{im} values are usually estimated by inverting (4.1) as follows:

$$\hat{f}_{im} \triangleq -\log\left(\frac{Y_{im} - r_{im}}{I_{im}}\right). \quad (4.8)$$

Component images are then estimated from the \hat{f}_{im} values in (4.8) by minimizing a PWLS cost function subject to box constraints on the elements of \mathbf{x} as follows:

$$\hat{\mathbf{x}} = \arg \min_{x_- \leq x_{lj} \leq x_+} \Psi(\mathbf{x}) \quad (4.9)$$

$$\Psi(\mathbf{x}) \triangleq \sum_{i=1}^{N_d} \sum_{m=1}^{M_0} \frac{1}{2} w_{im} \left(\hat{f}_{im} - f_{im}(\mathbf{s}_i(\mathbf{x})) \right)^2 + R(\mathbf{x}), \quad (4.10)$$

where $w_{im} = Y_{im}$ values are statistical weighting factors depending on Poisson likelihood, and $x_- \geq 0$ and x_+ are determined by the prior knowledge of ranges of material component values.

The edge-preserving regularizer term $R(\mathbf{x})$ is

$$R(\mathbf{x}) = \sum_{l=1}^{L_0} \beta_l R_l(\mathbf{x}_l), \quad (4.11)$$

where

$$R_l(\mathbf{x}_l) = \sum_{j=1}^{N_p} \sum_{k \in \mathcal{N}_{l_j}} \psi_l(x_{lj} - x_{lk}) \quad (4.12)$$

$$\psi_l(t) = \frac{\delta_l^2}{3} \left(\sqrt{1 + 3 \left(\frac{t}{\delta_l} \right)^2} - 1 \right), \quad (4.13)$$

where \mathcal{N}_{l_j} is some neighborhood of voxel x_{lj} . The regularization parameters β_l and δ_l can be chosen differently for different materials according to their properties.

4.3 Optimization Algorithm

Because the cost function $\Psi(\mathbf{x})$ in (4.10) is difficult to minimize directly, we apply optimization transfer principles to develop an algorithm that monotonically decreases $\Psi(\mathbf{x})$ each iteration. We ignore the regularizer term (4.11) hereafter since the challenging part is the nonlinear least-squares not the penalty function and the extension to PWLS is straightforward [34].

Define the vector beam-hardening function as

$$\mathbf{f}_i(\mathbf{s}_i) \triangleq (f_{i1}(\mathbf{s}_i), \dots, f_{iM_0}(\mathbf{s}_i)) \quad (4.14)$$

and rewrite the cost function (4.10) (ignore the regularizer term) as

$$\Psi(\mathbf{s}(\mathbf{x})) = \sum_{i=1}^{N_d} \Psi_i(\mathbf{s}_i(\mathbf{x})) = \sum_{i=1}^{N_d} \sum_{m=1}^{M_0} \Psi_{im}(\mathbf{s}_i(\mathbf{x})), \quad (4.15)$$

where

$$\Psi_i(\mathbf{s}_i) \triangleq \frac{1}{2} \left\| \hat{\mathbf{f}}_i - \mathbf{f}_i(\mathbf{s}_i) \right\|_{\mathbf{W}_i}^2 \quad (4.16)$$

$$\Psi_{im}(\mathbf{s}_i) \triangleq \frac{1}{2} w_{im} \left(\hat{f}_{im} - f_{im}(\mathbf{s}_i) \right)^2, \quad (4.17)$$

and $\mathbf{W}_i = \text{diag} \{w_{i1}, \dots, w_{iM_0}\}$. We first find a quadratic surrogate function for $\Psi_{im}(\mathbf{s}_i)$ since the cost function $\Psi_i(\mathbf{s}_i)$ is already additively separable in terms of \mathbf{s} , and then find a separable quadratic surrogate function of \mathbf{x} by applying De Pierro's additive convexity trick [2, 27].

4.3.1 Separable Quadratic Surrogate of Sinogram Vector

The separable quadratic surrogate function $\phi_1^{(n)}(\mathbf{s}_i)$ with respect to \mathbf{s}_i at the n th iteration that satisfies the surrogate conditions given in (2.67) has the following form

$$\phi_1^{(n)}(\mathbf{s}_i) = \sum_{i=1}^{N_d} \sum_{m=1}^{M_0} q_{im}^{(n)}(\mathbf{s}_i), \quad (4.18)$$

where

$$q_{im}^{(n)}(\mathbf{s}_i) \triangleq \Psi_{im}(\mathbf{s}_i) + (\mathbf{s}_i - \mathbf{s}_i^{(n)})' \nabla \Psi_{im}(\mathbf{s}_i^{(n)}) + \frac{1}{2} (\mathbf{s}_i - \mathbf{s}_i^{(n)})' \tilde{\mathbf{C}}_{im} (\mathbf{s}_i - \mathbf{s}_i^{(n)}), \quad (4.19)$$

where $\mathbf{s}_i^{(n)} \triangleq \mathbf{s}_i(\mathbf{x}^{(n)})$.

We use a bound on the Hessian of Ψ_{im} as the curvature $\tilde{\mathbf{C}}_{im}$. The gradient and Hessian of Ψ_{im} are

$$\nabla \Psi_{im}(\mathbf{s}_i) = -w_{im} \left(\hat{f}_{im} - f_{im}(\mathbf{s}_i) \right) \nabla f_{im}(\mathbf{s}_i) \quad (4.20)$$

$$\nabla^2 \Psi_{im}(\mathbf{s}_i) = H_1(\mathbf{s}_i) + H_2(\mathbf{s}_i), \quad (4.21)$$

where

$$H_1(\mathbf{s}_i) \triangleq w_{im} \nabla f_{im}(\mathbf{s}_i) \nabla f_{im}(\mathbf{s}_i) \quad (4.22)$$

$$H_2(\mathbf{s}_i) \triangleq w_{im} \left(\hat{f}_{im} - f_{im}(\mathbf{s}_i) \right) \left(-\nabla^2 f_{im}(\mathbf{s}_i) \right). \quad (4.23)$$

∇ and ∇ denote the column and row gradients respectively.

We ignore H_2 in (4.23) when finding an upper bound for Ψ_{im} since in practice we would hope that $\hat{f}_{im} \approx f_{im}(\mathbf{s}_i(\hat{\mathbf{x}}))$. It is stated in [111, p. 542] that the term $(\hat{f}_{im} - f_{im})$ should be the random measurement error of each point for a successful model. This error can have either sign, and should in general be uncorrelated with the model. Therefore, the second derivative terms H_2 tend to cancel out when summed over i . It is also mentioned that inclusion of the second-derivative term can in fact be destabilizing if the model fits badly or is contaminated by outlier points that are unlikely to be offset by compensating points of opposite sign. Thus, the goal simplifies to finding a positive non-negative matrix $\tilde{\mathbf{C}}_{im}$ such that $\tilde{\mathbf{C}}_{im} \succeq H_1(\mathbf{s}_i), \forall \mathbf{s}_i$. By calculation described in Appendix A.2, we choose

$$\tilde{\mathbf{C}}_{im} = \|\nabla f_{im}(\mathbf{0})\|^2 \mathbf{I}. \quad (4.24)$$

4.3.2 Separable Quadratic Surrogate of Image Vector

Define vector $\mathbf{x}_j \triangleq (x_{1j}, \dots, x_{L_0j})$, and rewrite the sinogram vector as $\mathbf{s}_i(\mathbf{x}) = \sum_{j=1}^{N_p} a_{ij} \mathbf{x}_j$. The surrogate function $\phi_1^{(n)}(\mathbf{s}_i)$ is a nonseparable quadratic function of \mathbf{x} . Nonseparable surrogates are inconvenient for simultaneous update algorithms and for enforcing the box constraint in (4.9). To derive a simple simultaneous update algorithm that is fully parallelizable and suitable for ordered-subsets implementation [2, 47], we find next a separable quadratic surrogate $\phi_2^{(n)}(\mathbf{x})$ by applying De Pierro's additive convexity trick [2, 27].

By derivations described in Appendix A.3, a separable quadratic surrogate that satisfies the conditions (2.67) has the form

$$\phi_2^{(n)}(\mathbf{x}) = \Psi(\mathbf{x}^{(n)}) + (\mathbf{x} - \mathbf{x}^{(n)})' \nabla \Psi(\mathbf{x}^{(n)}) + \frac{1}{2} (\mathbf{x} - \mathbf{x}^{(n)})' \text{diag}\{d_{l_j}^{(n)}\} (\mathbf{x} - \mathbf{x}^{(n)}), \quad (4.25)$$

where

$$d_{lj}^{(n)} = \sum_{i=1}^{N_d} \sum_{m=1}^{M_0} \frac{a_{ij}^2}{\pi_{ij}} \left[\tilde{\mathbf{C}}_{im} \right]_{ll}, \quad (4.26)$$

$$\pi_{ij} = \frac{|a_{ij}|}{a_i}, \quad a_i = \sum_{j=1}^{N_p} |a_{ij}|. \quad (4.27)$$

Because this $\phi_2^{(n)}(\mathbf{x})$ is a separable quadratic function, it is trivial to minimize, leading to the following parallelizable iteration:

$$\mathbf{x}^{(n+1)} = \left[\mathbf{x}^{(n)} - \text{diag} \left\{ \frac{1}{d_{lj}^{(n)}} \right\} \nabla \Psi(\mathbf{x}^{(n)}) \right]_+, \quad j = 1, \dots, N_p. \quad (4.28)$$

4.4 Simulation Results

To evaluate the proposed PWLS method, we reconstructed soft tissue and bone images of a NCAT chest phantom [129] from a single-energy CT scan using a simulated split and bow-tie filter. We generated the density images by simply adding the soft tissue and bone components, and synthesized the linear attenuation coefficient images at 511 keV by linearly combining the reconstructed components with their mass attenuation coefficients at that energy. We compared the soft tissue, bone, density and linear attenuation coefficient 511 keV reconstructed by the JS-FBP and proposed PWLS method. We also compared the result images of split and bow-tie filtration.

Fig. 4.3 shows the NCAT object used in the computer simulation. The units of \mathbf{x} are physical density (g/cm^3) and were assigned to 1.0 for soft tissue, 0.5 for lungs, 1.5 for spine and 2.0 for ribs. The lungs and soft tissue had the “soft tissue” characteristics shown in Fig. 2.3, and the spine and ribs had the “bone” characteristics of Fig. 2.3. The images were 512×512 and the pixel size was 1 mm. This chest phantom is unrealistic since there should be marrow (“soft tissue” characteristics) inside the cortical bones. Solid bones aggravate the beam-hardening effects, which helps to compare the extent that the proposed PWLS and JS-FBP method reduce these artifacts.

We simulated the geometry of a GE LightSpeed X-ray CT fan-beam system with an arc detector of 888 detector channels by 984 views over 360° . The size of each detector cell was 1.0239 mm. The source to detector distance was $D_{sd} = 949.075\text{mm}$, and the source to rotation center distance was $D_{s0} = 541\text{mm}$. We included a quarter detector offset to reduce aliasing. We used the distance-driven (DD) projector [26] to generate projections of the true object.

We simulated an incident spectrum of the X-ray tube voltage at 80 kVp. Fig. 4.2 shows the exiting spectra of this source spectrum screened by the split filter and body bow-tie filter described in Section 4.1.

We simulated noiseless measurements \bar{y}_{im} using (4.1) and the spectra shown in Fig. 4.2. To the noiseless measurements \bar{y}_{im} , we added Poisson distributed noise corresponding to 10^6 incident photons per ray to “soft” rays screened by the Al filter of the split filter. For “hard” rays and rays passing through the simulated bow-tie filter, we added Poisson noise corresponding to $10^6 I_i / I_j$ incident photons per ray where I_i and I_j denote the total intensity of the current ray and the “soft” ray respectively. Fig. 4.4 shows the logarithm sinogram estimates \hat{f}_{im} described in (4.8) as computed from the original noisy measurements Y_{im} . The visible straight line in the left sinogram was caused by the split filtration.

We first reconstructed FBP images using the water-correction method and the bone-correction JS method described in Section 2.1.4. Fig. 4.5 shows the FBP reconstructions and their absolute error images. The RMS errors were 0.14 g/cm^3 and $3.3 \times 10^{-2} \text{ g/cm}^3$ for the water and bone corrected FBP images using the split filtration, and 0.15 g/cm^3 and $3.1 \times 10^{-2} \text{ g/cm}^3$ respectively using the bow-tie filtration. The water corrected FBP had significant errors in the bone regions, while the JS-FBP greatly reduced them.

For the two cases of filtration, we separated the soft tissue shown in Fig. 4.6 and bone component shown in Fig. 4.7 from the JS-FBP reconstructions using a threshold of 1.2 g/cm^3 to initialize

the iterative optimization algorithm. We chose $\beta_1 = \beta_2 = 2^{-8}$ and $\delta_1 = \delta_2 = 0.01$ for soft tissue and bone components for both filtration methods. We ran 500 iterations of the optimization transfer algorithm described in Section 4.3 with 41 subsets to accelerate the “convergence” to the initialization. We enforced box constraints with a lower bound of 0 for both soft tissue and bone and upper bounds of 1.2 g/cm^3 and 2 g/cm^3 for soft tissue and bone respectively.

The soft tissue, bone and density images reconstructed by the JS-FBP and proposed PWLS method using split and bow-tie filtration were shown in Fig. 4.6, Fig. 4.7 and Fig. 4.8 respectively. Their RMS errors are summarized in Table 4.1. For both filtration methods, the RMS errors of the soft tissue and bone images reconstructed by the PWLS method were about 75% and 60% of those of the traditional non-iterative JS-FBP method respectively. For the density images, the RMS errors of the PWLS method using both filtration were about 40% of those by the JS-FBP method. The PWLS method removed the beam hardening artifacts much more effectively than the JS-FBP method which is known for mitigating these artifacts.

We synthesized the linear attenuation coefficient images at 511 keV by linearly combining the reconstructed components with their mass attenuation coefficients at that energy. Fig. 4.9 shows these attenuation images by the JS-FBP and PWLS method. The range of true linear attenuation coefficients was $[0 \text{ } 0.18] \text{ cm}^{-1}$. The reconstructed images were shown in a more sensitive viewing window ($[0.08, 0.12] \text{ cm}^{-1}$) to highlight the beam-hardening effects. The PWLS images had less beam-hardening artifacts. Table 4.1 also shows the RMS errors. The RMS errors of the PWLS method using both filtration were about 40% of those by the JS-FBP method.

Fig. 4.10 shows vertical profiles through the true and estimated soft tissue, bone components and the linear attenuation coefficient images at 511 keV by the JS-FBP and PWLS methods. The green line in the true linear attenuation image in Fig. 4.11 shows the profile locations. The proposed PWLS method produced lower noise than the JS-FBP method. The profiles showed negative bias

in the soft tissue regions and positive bias in the bone regions in the soft component, and similar opposite biases in the bone component. Profiles at other locations showed similar bias problem. We have also applied the proposed PWLS method to a more realistic simulated phantom that had marrow (soft tissue characteristics) inside cortical bones. This bias phenomenon remained. We found this bias problem is sensitive to regularizer coefficients β_l in (4.11) and edge-preserving parameters δ_l in (4.13). Increasing β_l and δ_l caused more bias but less noise, while decreasing them led to less bias but more noise. The choice of parameters for one material component influenced the reconstructed image of another component.

Fig. 4.11 compares the profiles of the PWLS results using the split and bow-tie filters. The split and bow-tie filtration had similar results, which was also demonstrated by reconstructed images in Fig. 4.5, Fig. 4.6, Fig. 4.7, Fig. 4.8, Fig. 4.9 and by the RMS errors in Table 4.1.

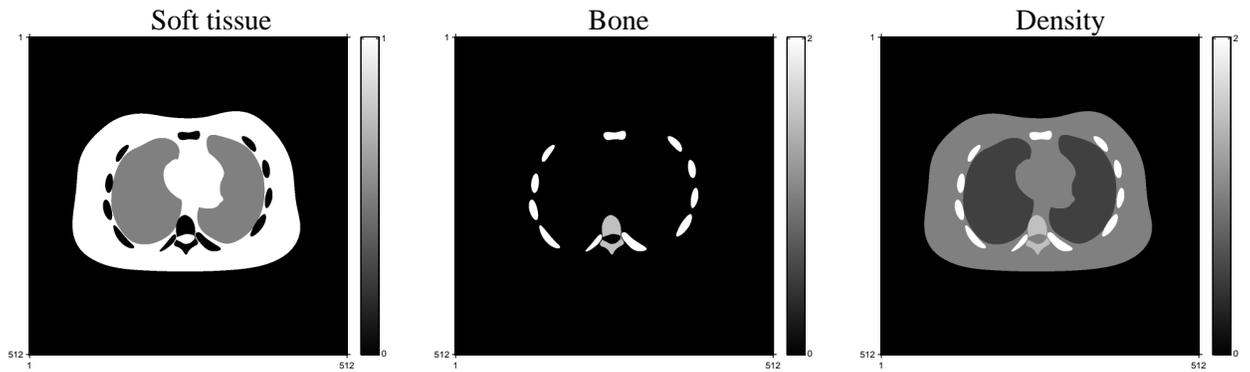


Figure 4.3: True soft tissue, bone and density map (soft tissue + bone) of a NCAT chest phantom.

Method	Split				Bow-tie			
	soft	bone	density	attenuation	soft	bone	density	attenuation
JS-FBP	4.0	3.4	3.2	3.0	4.0	3.3	3.0	2.8
PWLS	3.0	2.0	1.2	1.2	3.0	2.0	1.2	1.3

Table 4.1: RMS error comparison of reconstructed images by the JS-FBP and proposed PWLS method using split and bow-tie filtration. The units are 10^{-2}g/cm^{-3} for the soft tissue, bone and density images, and 10^{-3}cm^{-1} for the linear attenuation coefficient images at 511 keV.

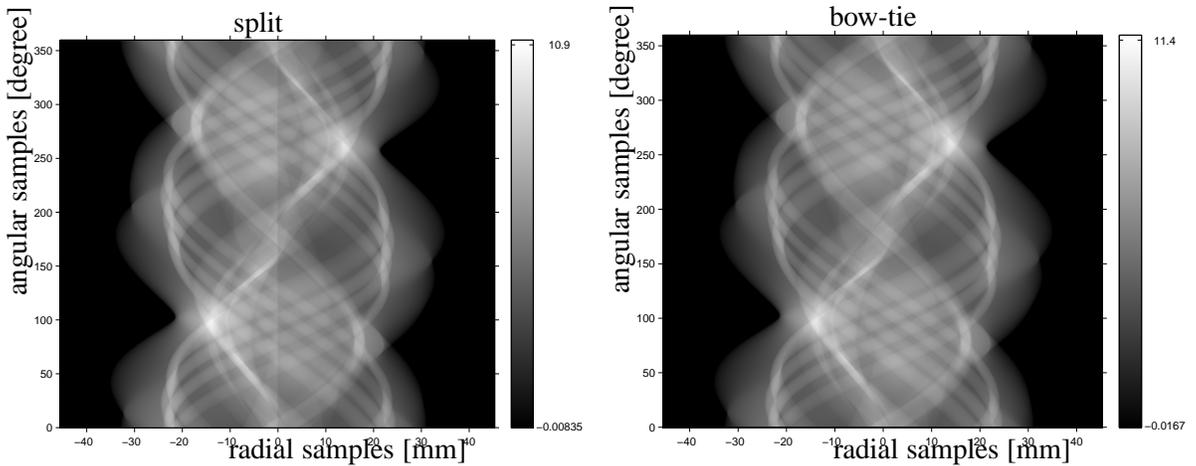


Figure 4.4: Logarithm sinogram estimates \hat{f}_{im} using the simulated split filter (*Left*) and bow-tie filter (*Right*).

4.5 Discussion and Conclusion

We have presented a statistical PWLS method for two-material decomposition from a single-energy CT scan acquired with a X-ray split filter or a bow-tie filter.

To monotonically decrease the PWLS cost function which was non-convex and non-linear, we presented an optimization transfer method with separable quadratic surrogates. To obtain a good initialization for the iterative optimization, we first reconstructed the bone-corrected FBP images using the Joseph and Spital method [57], and separated the soft and bone components by a threshold. We used the ordered subsets approach to accelerate the convergence to a good local minimum. We enforced box constraints on both soft tissue and bone estimates, but found the reconstruction results were insensitive to the bounds of box constraints.

Experiments showed that the proposed PWLS method reconstructed soft tissue and bone components with 25% and 40% lower RMS errors respectively, and reconstructed density (the sum of soft tissue and bone) images and linear attenuation coefficient images at 511 Kev with 40% lower RMS errors than the JS-FBP method. The PWLS images had fewer beam-hardening artifacts and

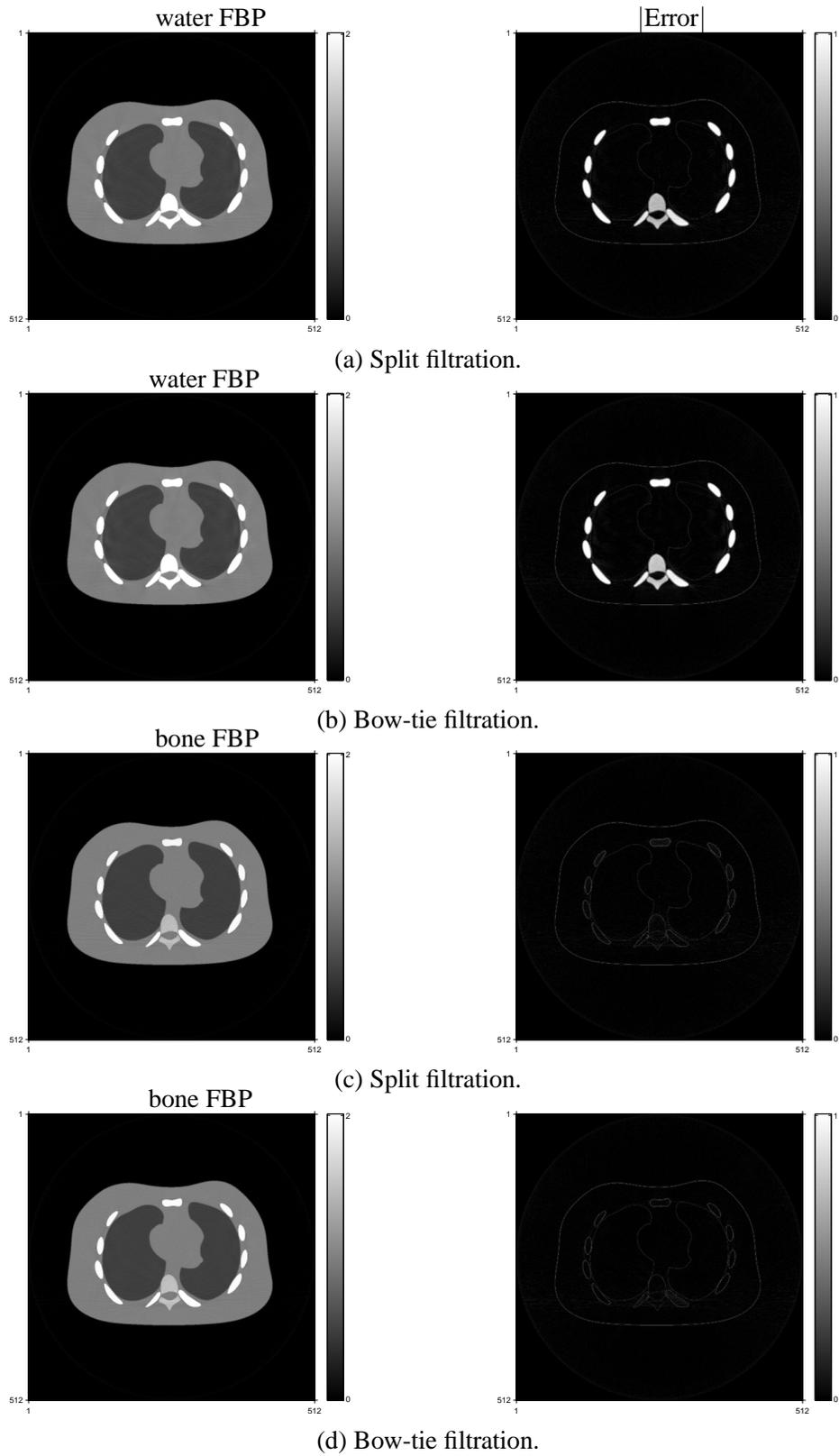


Figure 4.5: Water and bone corrected FBP and their absolute error images using split and bow-tie filtration.

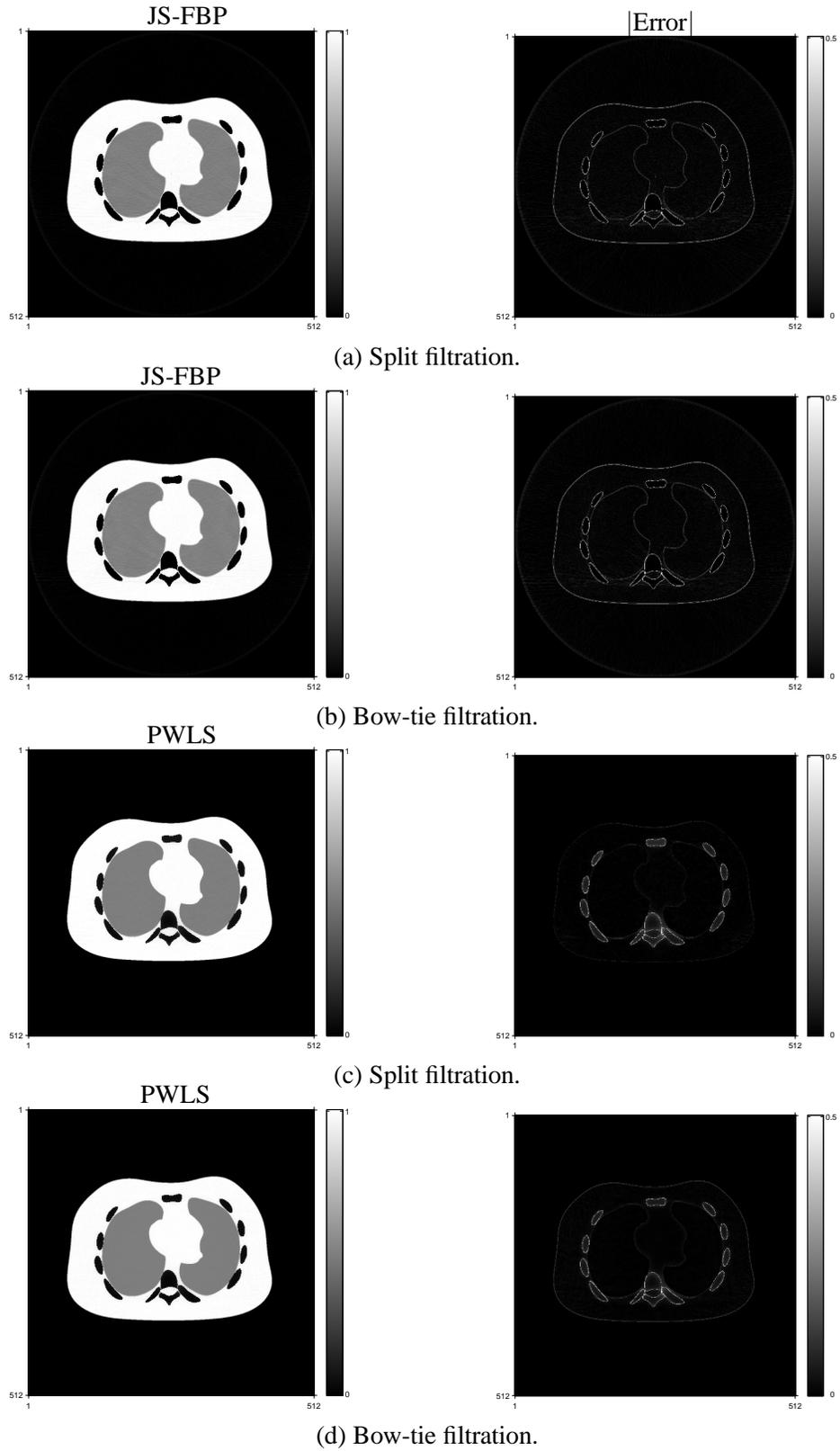


Figure 4.6: Soft tissue images and their absolute errors reconstructed by the JS-FBP and proposed PWLS method using split and bow-tie filtration.

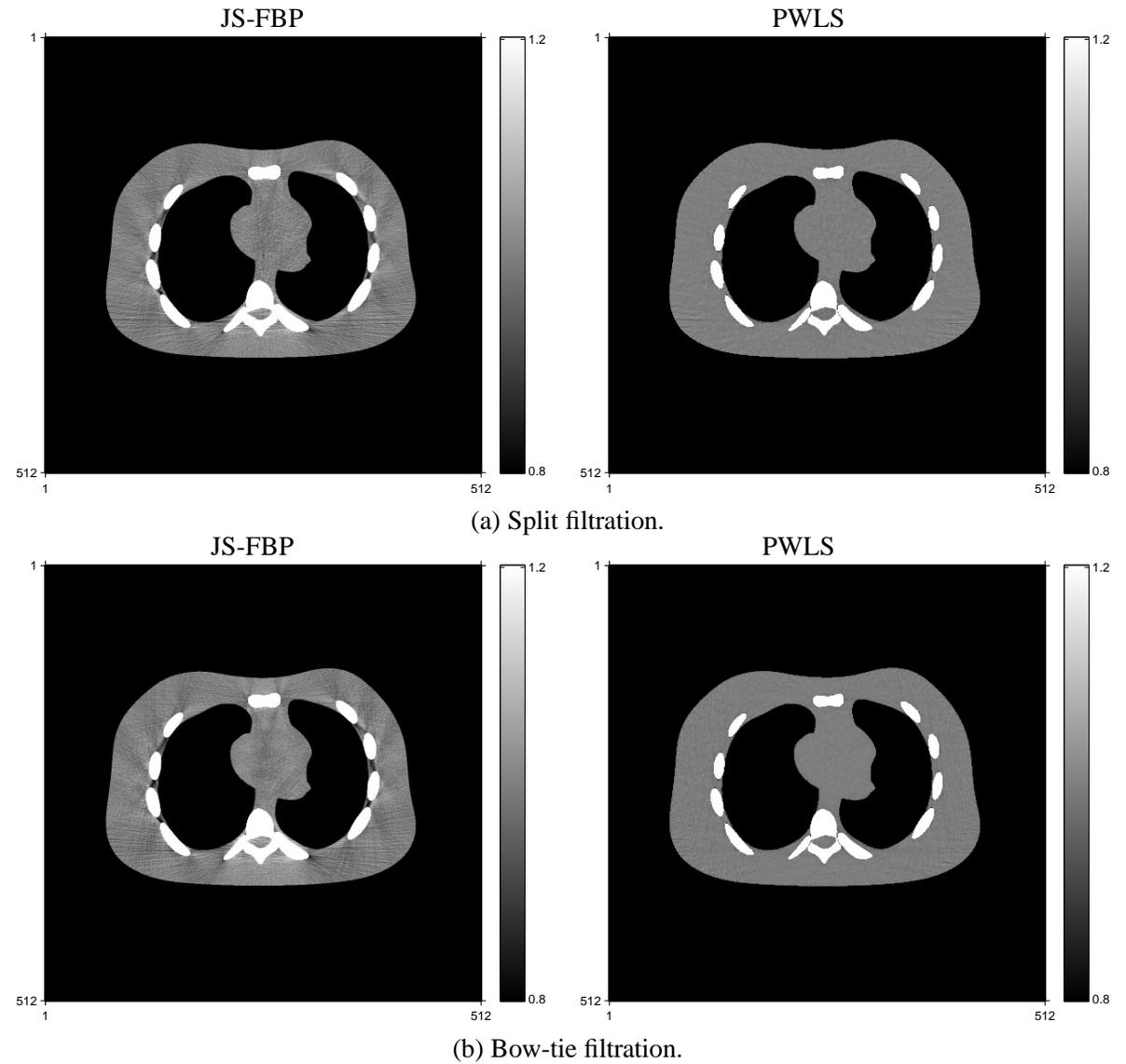


Figure 4.8: Density maps (the sum of soft tissue and bone images) reconstructed by the JS-FBP and PWLS method using split and bow-tie filtration. The images are shown in a more sensitive viewing window ($[0.8, 1.2]$ g/cm³).

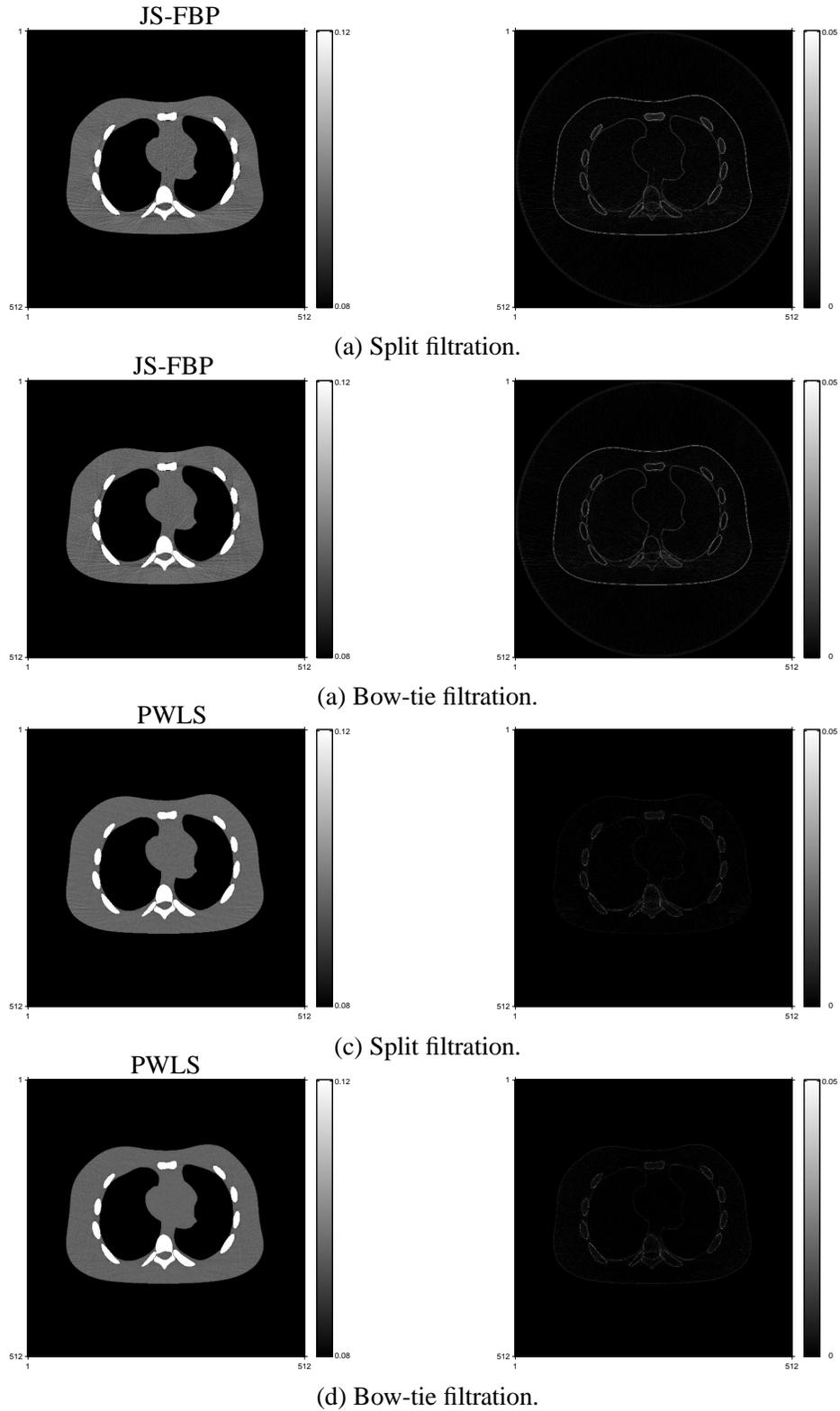


Figure 4.9: Linear attenuation coefficient images at 511 keV reconstructed by the JS-FBP and PWLS method using split and bow-tie filtration. The range of true linear attenuation coefficients was $[0 \ 0.18] \text{ cm}^{-1}$. The images are shown in a more sensitive viewing window ($[0.08, 0.12] \text{ cm}^{-1}$) to highlight the beam-hardening effects.

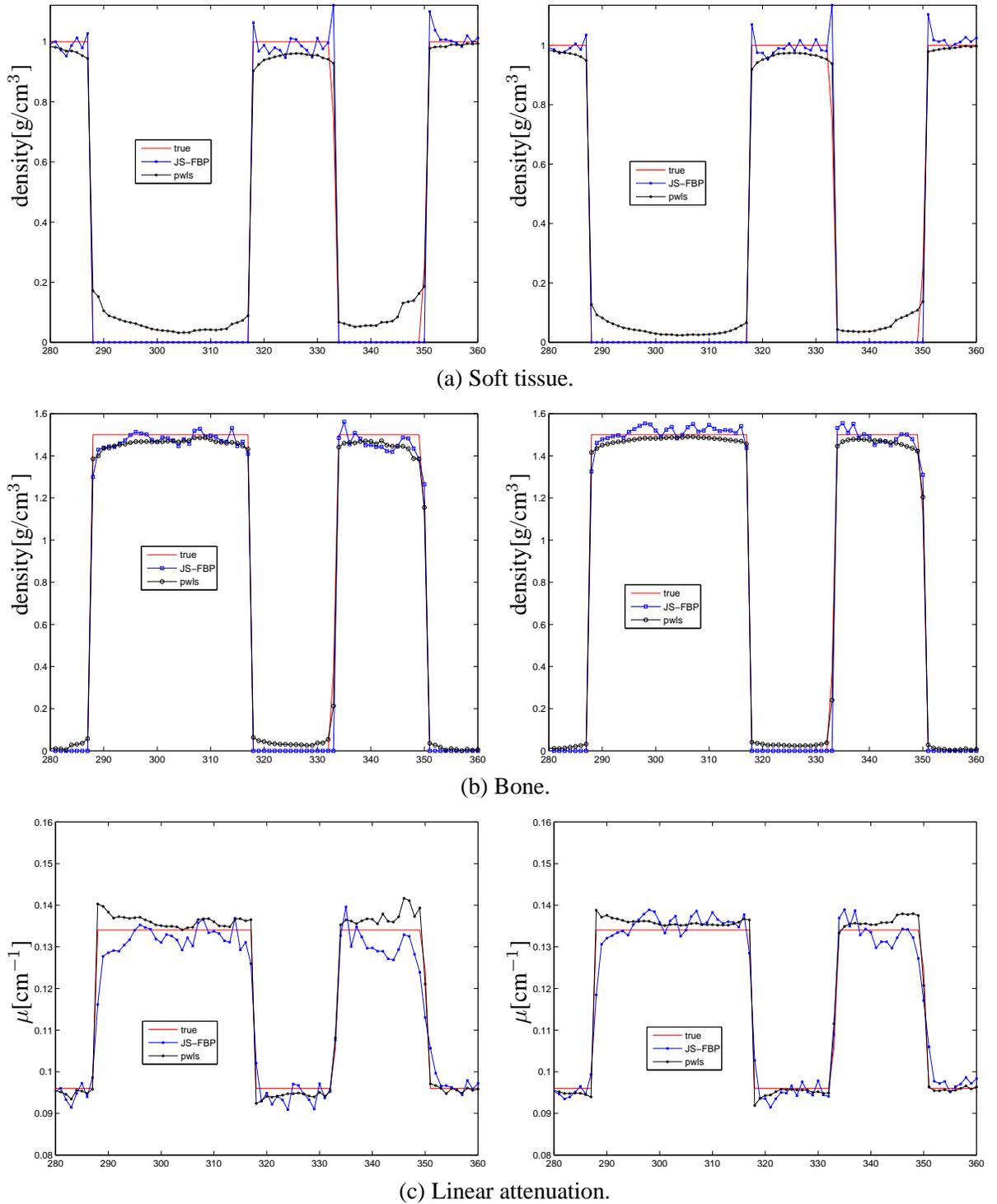


Figure 4.10: Vertical profiles through the true and estimated soft tissue, bone components and the linear attenuation coefficient images by the JS-FBP and PWLS method using split (*left*) and bow-tie (*right*) filtration. The green line in the true density image in Fig. 4.11 shows profile locations.

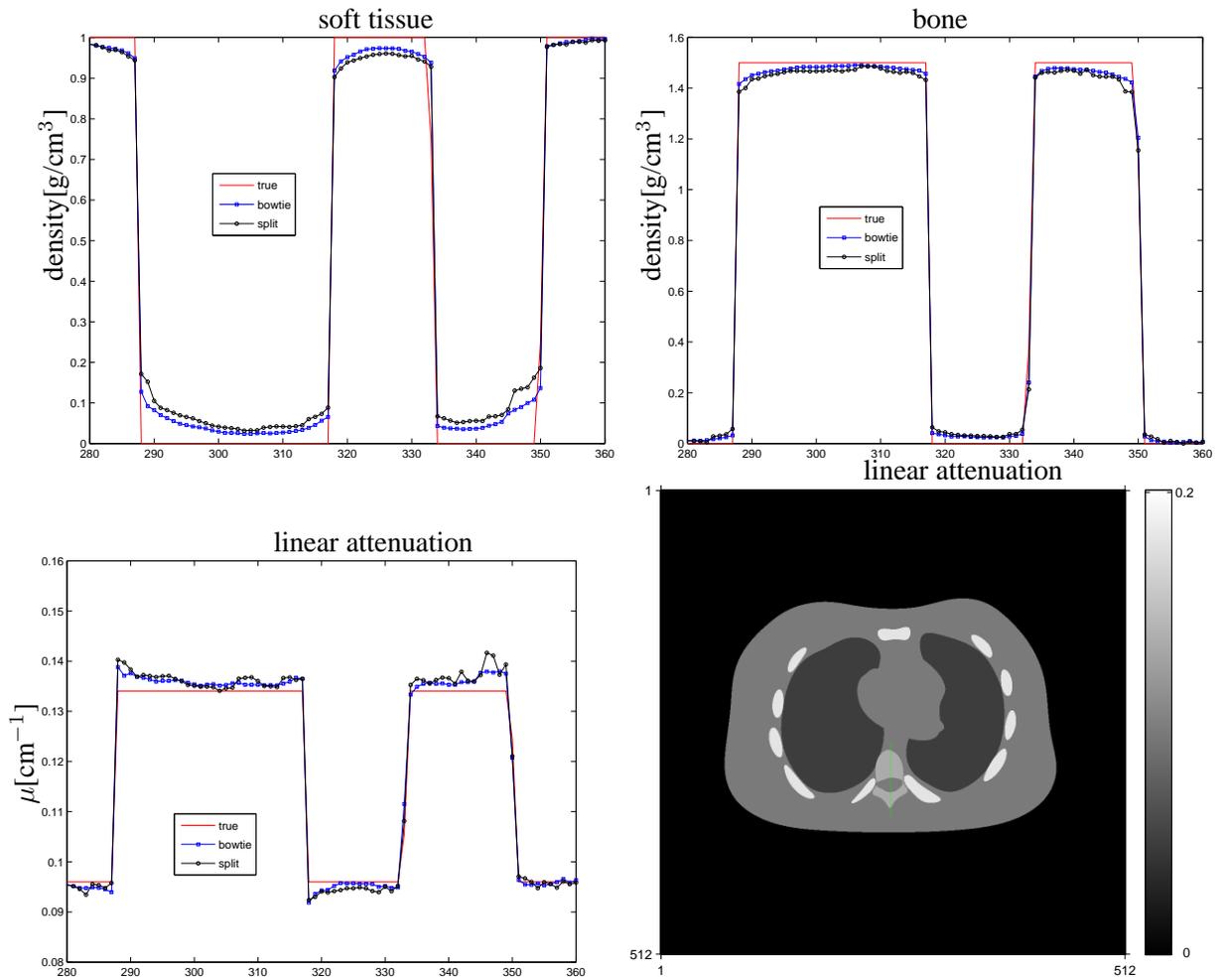


Figure 4.11: Vertical profiles of PWLS method using the simulated bow-tie and split filter. The green line in the true linear attenuation image at 511 keV shows the location of profiles in this figure and Fig. 4.10.

lower noise than the JS-FBP method. The bow-tie filtration produced similar results as the split filtration. We did not compare the spatial resolutions of reconstructions by the JS-FBP and PWLS methods, because it is not easy to evaluate the spatial resolution properties of reconstruction methods with edge-preserving regularization [33, 136, 137]. However, the profiles in Fig. 4.10 looks similar.

The role of the split or bow-tie filter is to create differences between incident spectra to reduce the ill-conditioning of the problem of decomposing two materials from one single-energy CT scan. We simulated a split filter with 2 mm aluminum across half the fan beam and 200 μm molybdenum of the other half [146]. The simulated body bow-tie filter contains aluminum, graphite, copper and titanium. Performed experiments showed our simulated filters provide sufficient spectra differences for the PWLS method to reconstruct soft tissue and bone components. We believe other choices of materials and thickness for the split and bow-tie filters or other types of differential filtration could lead to similar or even better results for the proposed PWLS method. Optimizing the materials and thickness or the types of differential filtration is not our focus. We simulated an incident spectrum for an 80 kVp of the X-ray tube voltage. One could choose other X-ray tube voltages. Different voltage settings could affect the degree of spectral separation produced by split and bow-tie filters, and consequently could change the reconstructed component images. The radiation dose should be carefully adjusted according to the X-ray tube voltage.

A fundamental limitation of using differential filtration with a single-energy scan is the inevitable overlap of the filtered spectra (see Fig. 4.2). We expect that DECT methods are superior from the standpoint of spectral differences that may lead to better reconstructed component image quality. However, the differential filtration requires only the attachment and alignment of the metal filters on the X-ray tube housing. Another problem of using differential filtration is to precisely align the filters and rotational center in order to expose the isocenter area to different spectra

which helps reconstructing component images in that area. We simulated a 2D fan-beam system to evaluate the proposed PWLS method. For this geometry, each path in the patient cross section is measured with two different beam filtration created by a split filter during a single 360° rotation scan. However, for typical clinical CT systems, such as axial and helical cone beam geometries, most tilted paths in the patient are not passed by two different incident spectra. This problem remains to be addressed for practical use of split filters. We also expect the results could be sensitive to model mismatch, such as Compton scatter or imperfect spectral models.

The PWLS cost function has four parameters, two regularizer coefficients β_l in (4.11) and two edge-preserving parameters δ_l in (4.13). We found slight negative bias in the soft tissue regions and positive bias in the bone regions in the soft component, and similar opposite biases in the bone component. This bias problem is sensitive to these parameters. Increasing them caused more bias but less noise, while decreasing them led to less bias but more noise. The choice of parameters for one material component influenced the reconstructed image of another component. Huh and Fessler [49] used a material-cross penalty for DECT reconstruction. This penalty used the prior knowledge that different component images have similar edge area. One could also consider this basis material reconstruction problem as a classification problem, since most pixels either have soft tissue characteristics or bone characteristics [29]. Penalty functions could be designed based on this prior knowledge. Choosing regularizers for the PWLS method and optimizing its parameters need further investigation.

Two component images are not enough for some medical applications [78, 89, 96]. When quantifying the concentration of iron in a fatty liver, images of three constitute materials, iron, fat and tissue, are required. For the purpose of radiotherapy, in addition to soft-tissue and bone it is also better to know the distribution a third material, such as calcium, metal (gold) or iodine. An interesting future research would be to extend this PWLS method to three-material decomposition

from dual-energy CT scan (such as fast kVp-switching and dual-source CT) acquired with a split or bow-tie filter. The use of filtration for DECT creates differential spectra to the original spectra at two source voltage settings, which would provide additional information for separating three materials.

CHAPTER 5

Accuracy Estimation for Projection-to-Volume Targeting During Rotational Therapy: A Feasibility Study ¹

The rapid adoption of two technologies, arc therapies [105, 159] and cone-beam CT [21, 54, 55, 98], have brought to the forefront a number of investigations about optimizing the use of projection radiographs and reference volumes for targeting treatment. While a significant number of investigations are ongoing into optimizing reconstruction and use of cone beam CT, a smaller but highly relevant path of investigation is similarly being pursued in using subsets of projections from a rotational series for alignment and reconstruction.

Such experiments fall into two primary classes of operation. The first involves reconstructing volumetric images from subsets of projections acquired over rotational arcs of various lengths. For longer arcs, these backprojections yield volumetric images with fairly uniform resolution, while for shorter (typically 45-90 degree) arcs, the reconstruction yields volumes with spatially varying resolution. Such reconstructions generally are termed tomosynthesis, and have been evaluated from both kilovoltage as well as megavoltage projections for use in image guided radiotherapy [41, 67, 124, 150]. There is a special case of tomosynthesis using a fixed set of projections

¹This chapter is based on material from [82, 83].

(more complex than an arc) [90]. These reconstructed volumes are typically aligned directly to the reference CT scan.

A different approach involves relating the new projections directly to the reference CT without reconstruction. Such projection-to-volume alignment experiments have been performed for some time [4, 63]. More recently, the series of projections acquired during arc rotations has been used to estimate motion and deformation parameters [77, 95, 106, 114, 115, 155, 158, 163, 164].

This area of research is highly promising. As the amount of information needed to estimate the position, pose, configuration, and finally motion of a patient is reduced, the temporal resolution of updates to the patient state improves. In this investigation, methods for estimating the local configuration of a lung tumor are tested, and an evaluation is made of the influence of rotational arc length on maximum achievable accuracy.

5.1 Materials and Methods

5.1.1 Projection-to-Volume Registration

We assume a static reference volume f^{ref} , such as a breath-hold planning CT, is available, and we record a sequence of cone-beam projection views of a limited-angle scan, denoted as y_n for $n = 1, \dots, N$ where N is the number of projection views. The current experiment assumes that movement during the limited projection arc is negligible. One can imagine that a continuous relationship exists between temporal resolution and spatial accuracy. The current investigation focuses on the spatial accuracy under ideal conditions (no movement between projections). While one can assume that movement during rotation would decrease the accuracy of measurements, it is our expectation that further studies will be able to extend prior models for estimating motion during rotational arcs (*e.g.*, Refs. [77, 164]), but that it is unlikely that such methods can recover spatial information at accuracies higher than the current motion-free estimations are capable of.

We estimate the local configuration of a lung tumor using two sets of data, the reference volume \mathbf{f}^{ref} and projection views \mathbf{y}_n .

A nonrigid model is suitable to describe changes in thorax configuration during breathing or as varying between breath held states. We use the cubic B-spline nonrigid motion model given in (2.55), and assume a target image \mathbf{f}^{tar} is deformable from \mathbf{f}^{ref} (See (2.62)). We estimate the deformation parameters $\boldsymbol{\alpha}$ by minimizing the following regularized cost function:

$$\begin{aligned}\hat{\boldsymbol{\alpha}} &= \arg \min_{\boldsymbol{\alpha}} \Psi(\boldsymbol{\alpha}) \\ \Psi(\boldsymbol{\alpha}) &= L(\boldsymbol{\alpha}) + \gamma \mathbf{R}(\boldsymbol{\alpha}),\end{aligned}\tag{5.1}$$

where $L(\boldsymbol{\alpha})$ is the data fidelity term, $\mathbf{R}(\boldsymbol{\alpha})$ is the regularization function, and scalar γ controls the trade-off between them.

We focus on the least-squared error metric, because all the investigations in this paper are based on the same patient and imaging modality. The metric is expressed as,

$$L(\boldsymbol{\alpha}) = \frac{1}{2} \sum_{n=1}^N \|\mathbf{y}_n - \mathbf{A}_{\varphi_n} \mathbf{T}(\boldsymbol{\alpha}) \mathbf{f}^{\text{ref}}\|^2,\tag{5.2}$$

where \mathbf{A}_{φ_n} denotes the system matrix with size of $N_d \times N_p$ at projection angle φ_n that is the angle of the source point counter-clockwise from the y axis (see Fig. 5.2), $\mathbf{y}_n \triangleq (y_{n,1}, \dots, y_{n,l}, \dots, y_{n,N_d})$ is the n th cone-beam CT projection view, $\mathbf{T}(\boldsymbol{\alpha})$ is given in (2.62) and (2.63). In practice \mathbf{y}_n is estimated from the transmitted intensity \mathbf{Y}_n which is degraded by noise that dominated by the Poisson effect [1]. For simplicity, assuming a monoenergetic model and ignoring the background signals such as Compton scatter, dark current and noise, we describe the Poisson statistics as follows,

$$Y_{n,l} \sim \text{Poisson}\{I_{n,l} e^{-y_{n,l}}\},\tag{5.3}$$

where $I_{n,l}$ denotes the incident intensity that incorporates the source spectrum and the detector

gain. The projection view \mathbf{y}_n is estimated from \mathbf{Y}_n as follows,

$$\hat{y}_{n,l} = \log\left(\frac{I_{n,l}}{Y_{n,l}}\right). \quad (5.4)$$

To encourage local invertibility and smoothness of changes due to local respiratory motion and its variations, we adopted a simple regularizer proposed by Chun *et al.* [23], which is introduced in Section 2.2.3. In this investigation we used the same parameter settings as in Section IV-D of Ref. [23].

For optimization of the the cost function (5.1), a conjugate gradient (CG) method was used. The line search step size was determined by one step of Newton’s method. To avoid local minima and accelerate the optimization procedure, we applied a 4-level multi-resolution scheme [152]. We ran 100 iterations of conjugate gradient optimization for the first three levels of resolution and 150 iterations for the finest resolution. Large number of iterations were used to ensure convergence. Since this paper is a study of estimation accuracy, the computation time is not the major concern.

5.1.2 Investigating the Influence of Rotational Arc Length

Experimental Setup

We acquired three 3D thorax volumes, two at different inhale states and one at exhale state, from a lung cancer patient by breath-held diagnostic CT. One inhale CT volume was chosen as the reference \mathbf{f}^{ref} , while the other inhale and the exhale volume were treated as different targets for evaluation. The volume size was $512 \times 512 \times 96$ with spacing of $0.9375 \times 0.9375 \times 3\text{mm}^3$ in the x , y and z direction respectively.

Two deformation maps were obtained by regularized B-spline nonrigid registration [23] between reference and each of the target volumes. The B-spline control knots were placed uniformly in CT volumes with a spacing of $8 \times 8 \times 4$ voxels. Two synthetic CT volumes (see Fig. 5.1), one at

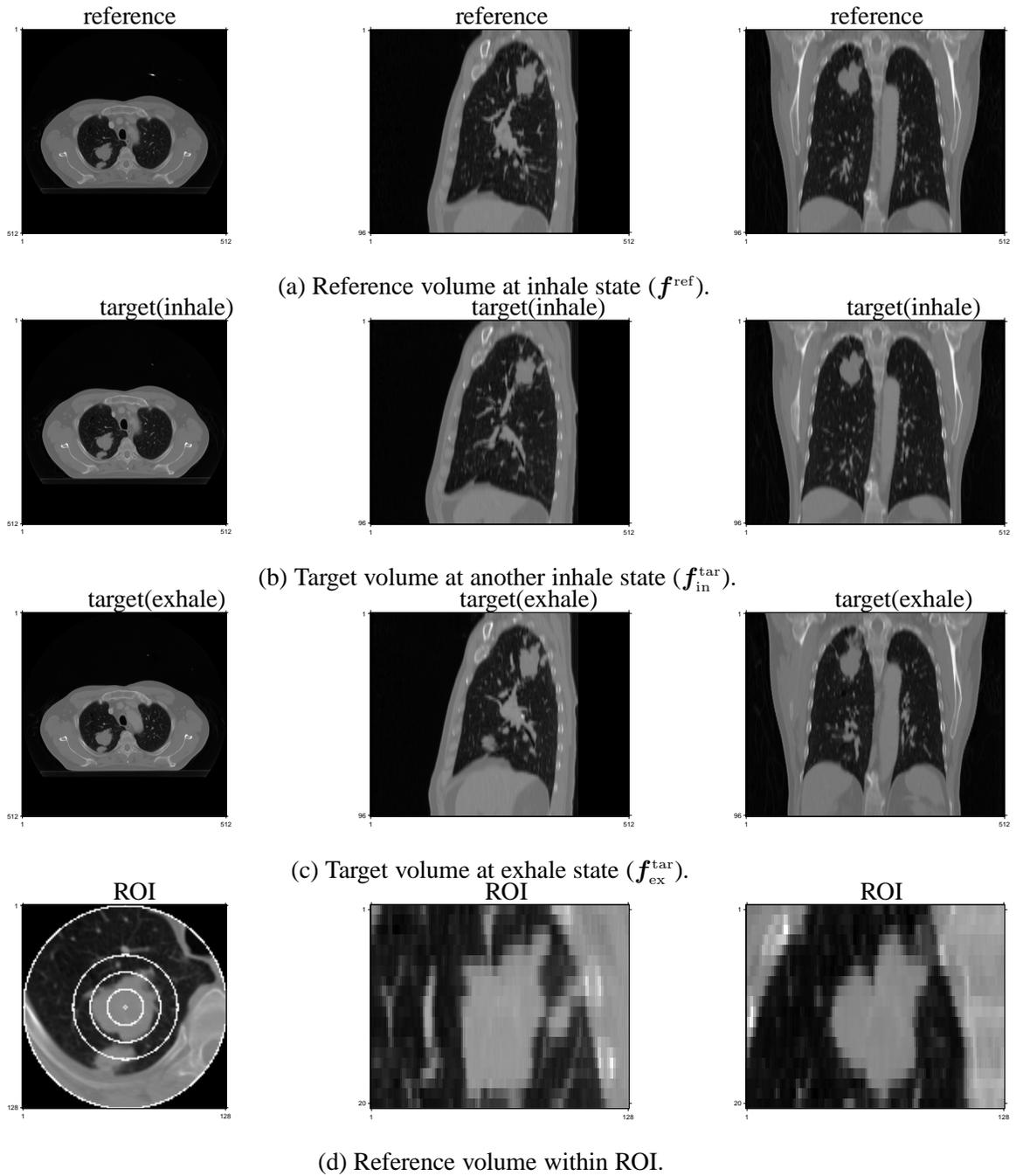


Figure 5.1: Axial (*left*), sagittal (*center*) and coronal views (*right*) of reference, target volumes and reference volume within a ROI overlaid with example radial annuli for analysis.

inhale and another at exhale, were obtained by warping the reference volume with corresponding estimated deformation maps in a fashion similar to that applied in previous research (Ref. [164]). These two synthetic CT volumes were used as new targets, denoted as $f_{\text{in}}^{\text{tar}}$ and $f_{\text{ex}}^{\text{tar}}$, for subsequent investigations of influence of rotational arc length on accuracy of estimated local deformation around the tumor. The B-spline control knots were also placed uniformly in CT volumes with a spacing of $8 \times 8 \times 4$ voxels, implying $64 \times 64 \times 24 = 98304$ unknown deformation parameters to be estimated.

We simulated an axial cone-beam flat-detector X-ray CT system with a detector size of 512×512 cells spaced by $1 \times 1\text{mm}^2$. The source to detector distance is 1500mm, and the source to rotation center distance is 1000mm.

Since our focus is the local configuration of a lung tumor (see Fig. 5.1), we set the tumor center in the reference volume as the rotation center to make sure the local areas around it in the target volumes are always covered by X-rays emitting from the source at any projection angle. As a result, other structures away from the tumor, such as the scapular bones in the lower right or left sides (see Fig. 5.1), might be truncated at certain projection angles.

Investigation Design

Typically the X-ray source rotates around the patient, and 360° of projection views are acquired to perform projection-to-volume alignment. In this paper, we investigate deformation estimation accuracy with limited-angle scans (see Fig. 5.2). In addition, the angular center may affect the estimation accuracy. Without loss of generality, we chose three angular centers (0° , 45° and 90°) and six angular ranges (12° , 18° , 24° , 36° , 60° and 90°) with 2° angular spacing between projections. (We examined smaller angular spacings of 1° and 0.5° , and found performances very similar to that of 2°). Using the case of 0° center and 12° range as an example, we performed projection-to-

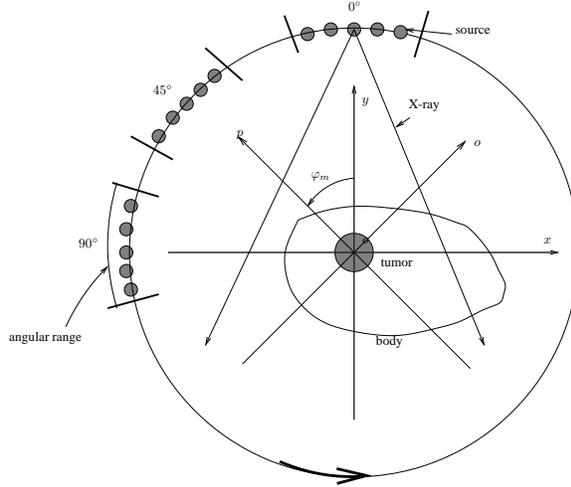


Figure 5.2: Schematic diagram of limited-angle scans centered at 0° , 45° and 90° . p and o are the axes along and orthogonal to the ray connecting the X-ray source and the detector center for the limited-angle scan centered at 45° .

volume alignment between projection views from 7 angles (-6° , -4° , -2° , 0° , 2° , 4° and 6°) and the reference volume. Alignments on angular ranges of 360° and 180° ($[-90^\circ, 90^\circ]$) were investigated too.

We also studied the influence of extent of deformation. We considered the deformation between two different inhale volumes as small, and the deformation between inhale and exhale volumes large. We called the former as *small deformation case* and the latter *large deformation case*. The experiments were executed on both cases.

In summary, we studied a total of 40 registration examinations. For each (small and large) deformation case we tested 20 examinations that include the 360° and 180° scan and another 18 limited-angle scans (6 angular ranges at 3 centers).

5.2 Results and Discussion

In this section, we summarize the error of tumor center shift and deformation estimation within a region of interest (ROI) around the tumor. We computed the error using the true deformation

maps between the reference and synthetic target volumes in comparison to the experimental alignment estimates from various angular centers and ranges.

5.2.1 Noiseless Projection Views

We generated noiseless projection views of target volumes ($f_{\text{in}}^{\text{tar}}$ and $f_{\text{ex}}^{\text{tar}}$) at certain angles using a distance-driven (DD) forward projector [26] for A_φ . The projection-to-volume registration described in Section 5.1.1 was performed between these generated projection views and the reference volume.

Accuracy of Tumor Center Shift

We calculated the absolute error of estimated tumor center shift in C direction by

$$E_0^C = \left| \hat{\mathcal{D}}_C(\vec{r}_0; \hat{\alpha}) - \mathcal{D}_C(\vec{r}_0; \alpha) \right|, \quad (5.5)$$

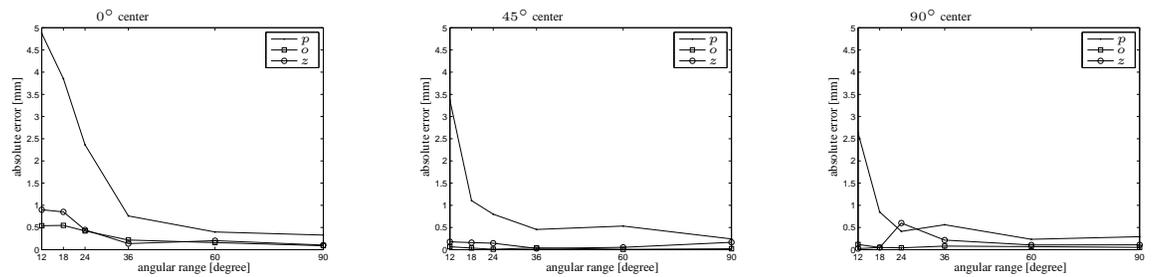
where $\hat{\alpha}$ denotes the estimated motion parameters, $\hat{\mathcal{D}}_C(\vec{r}_0; \hat{\alpha})$ denotes the estimated tumor center shift in C direction, $\mathcal{D}_C(\vec{r}_0; \alpha)$ denotes the true tumor center shift in C direction and \vec{r}_0 denotes the coordinates of the tumor center.

The true shifts of tumor center are -2.21mm , -2.46mm and 0.56mm in x , y and z direction respectively for the small deformation case, and -0.92mm , 6.17mm and 1.53mm in x , y and z direction respectively for the large deformation. Table 5.1 shows the absolute errors of tumor center shift when the angular ranges are 360° and 180° . The absolute errors are all below 0.1mm , except the error of 0.13mm in z direction for the large deformation case. Since the true shift is 1.53mm for this case and the spacing is 3mm in z direction, this 0.13mm error is small.

The projection views are approximately line integrals along rays passing from the X-ray source to the detector cells [26]. With limited-angle scans, the information about deformation along the projection direction is limited. Realizing this property, we used a 3D coordinate system with new

	small deformation			large deformation		
direction	x	y	z	x	y	z
360°	0.06	0.08	0.03	0.09	0.06	0.08
180°	0.08	0.07	0.001	0.08	0.05	0.13

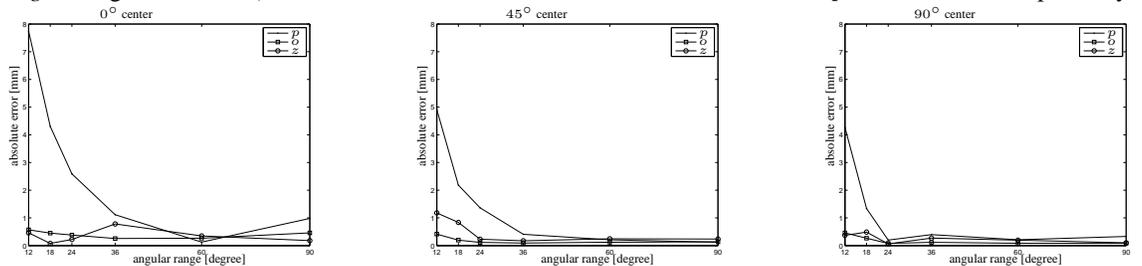
Table 5.1: Absolute errors (mm) for angular ranges of 360° and 180° ($[-90^\circ, 90^\circ]$).



(a) Small deformation case.

The true shift is 0.56mm in z direction.

Left: Angular center is $\beta = 0^\circ$. The true shifts are -2.46mm and -2.21mm in p and o direction respectively.
Center: Angular center is $\beta = 45^\circ$. The true shifts are -0.18mm and -3.30mm in p and o direction respectively.
Right: Angular center is $\beta = 90^\circ$. The true shifts are -2.21mm and -2.46mm in p and o direction respectively.



(b) Large deformation case.

The true shift is 1.53mm in z direction.

Left: Angular center is $\beta = 0^\circ$. The true shifts are 6.17mm and 0.92mm in p and o direction respectively.
Center: Angular center is $\beta = 45^\circ$. The true shifts are 5.01mm and 3.71mm in p and o direction respectively.
Right: Angular center is $\beta = 90^\circ$. The true shifts are -0.92mm and 6.17mm in p and o direction respectively.

Figure 5.3: Absolute errors (mm) of tumor center shift.

axes to evaluate estimation accuracy. We rotated the 3D coordinate system around z by the central projection angle φ_0 , and denoted the axes as p , o and z where p and o are the axes along (parallel to) and orthogonal to the ray connecting the X-ray source and the detector center respectively. Fig. 5.2 shows the p and o directions when $\varphi_0 = 45^\circ$. The corresponding coordinates on the p and o axes are:

$$\begin{aligned} p &= -x \sin(\varphi_0) + y \cos(\varphi_0) \\ o &= x \cos(\varphi_0) + y \sin(\varphi_0). \end{aligned} \quad (5.6)$$

Fig. 5.3 shows the absolute errors of tumor center shift with limited-angle scans centered at 0° , 45° and 90° for both the small and large deformation cases. The error in the p direction is bigger than that in the o and z direction because only limited shape information can be extracted from projection views along the projection direction. The errors of the 0° center scans are larger than those of 45° and 90° center scans. When angular range is smaller than 36° , the estimation accuracy improves quickly with the increase of angular range, and the estimation errors of the large deformation case are slightly higher than those of the small deformation case. When the angular range exceeds 36° , the errors are within 1mm for the 0° center scans, and within 0.5mm for others.

Deformation Accuracy within A ROI

Since the goal is to study estimation accuracy of the local configuration of a lung tumor, we chose a region of interest (ROI) centered at the tumor center and evaluated deformation estimation accuracy within this ROI. The ROI is a cylinder with height and radius of 6cm, *i.e.*, a diameter of 128 voxels and height of 20 voxels. The tumor center is also the rotation center, which guarantees the ROI is covered by the field of view (FOV) of radiation at any projection angle. Fig. 5.1 (d)

	small deformation						large deformation					
	angular range 180°			angular range 360°			angular range 180°			angular range 360°		
direction	x	y	z	x	y	z	x	y	z	x	y	z
mean	0.08	0.07	0.09	0.05	0.05	0.07	0.06	0.07	0.08	0.05	0.06	0.06
max	1.06	1.38	1.83	0.84	0.77	1.06	0.54	0.75	1.74	0.44	0.75	0.62
σ	0.10	0.10	0.11	0.06	0.06	0.08	0.07	0.07	0.09	0.05	0.07	0.06

Table 5.2: Mean, maximum and standard deviation σ of absolute errors of estimated displacements with angular ranges of 180° ($[-90^\circ, 90^\circ]$) and 360°.

shows axial, sagittal and coronal views of the reference volume within ROI.

We evaluated absolute errors of estimated deformation of points in a set of S , such as the ROI, by mean E_1^C , maximum E_2^C and standard deviation E_3^C in C direction as follows

$$\begin{aligned}
E_1^C &= \frac{1}{|S|} \sum_{j \in S} \left| \hat{\mathcal{D}}_C(\vec{r}_j; \hat{\alpha}) - \mathcal{D}_C(\vec{r}_j; \alpha) \right| \\
E_2^C &= \max_{j \in S} \left| \hat{\mathcal{D}}_C(\vec{r}_j; \hat{\alpha}) - \mathcal{D}_C(\vec{r}_j; \alpha) \right| \\
E_3^C &= \sqrt{\frac{1}{|S| - 1} \sum_{j \in S} \left(\left| \hat{\mathcal{D}}_C(\vec{r}_j; \hat{\alpha}) - \mathcal{D}_C(\vec{r}_j; \alpha) \right| - E_1^C \right)^2}, \quad (5.7)
\end{aligned}$$

where $\hat{\alpha}$ denotes the estimated motion parameters, $\hat{\mathcal{D}}_C(\vec{r}_j; \hat{\alpha})$ and $\mathcal{D}_C(\vec{r}_j; \alpha)$ denote the estimated and true displacement at the j th point in S in the C direction respectively, and \vec{r}_j denotes the coordinates of the j th point.

Table 5.2 shows the mean, max and standard deviation of estimated deformation for the angular ranges of 360° and 180°. The errors are very small, which demonstrates that the projection-to-volume method described in Section 5.1.1 works well.

Fig. 5.4, Fig. 5.5 and Fig. 5.6 show the mean, max and standard deviation of estimated deformation of all the studied limited-angle scan cases. In general, values of these measures decrease with increasing angular range. For 45° and 90° center cases, the errors and standard deviation in the p direction are bigger than those in the o and z direction and the errors of the large deformation case are slightly bigger than those of the small deformation case. The mean absolute error is

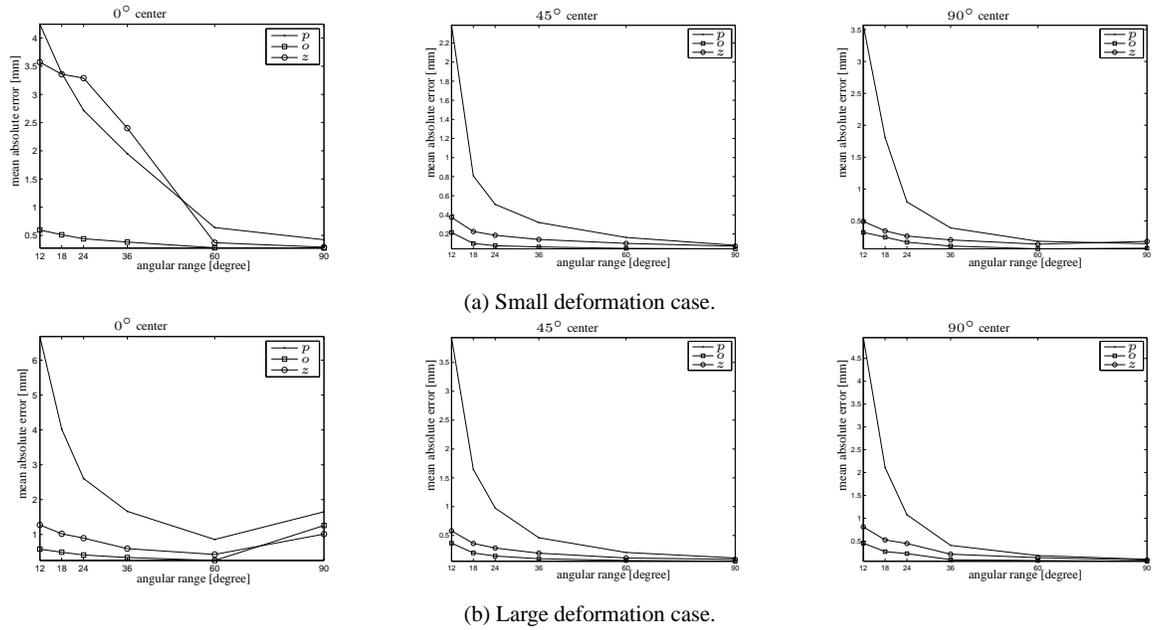


Figure 5.4: Mean absolute errors (mm) of estimated deformation within ROI. *Left:* Angular center is $\beta = 0^\circ$. *Center:* Angular center is $\beta = 45^\circ$. *Right:* Angular center is $\beta = 90^\circ$.

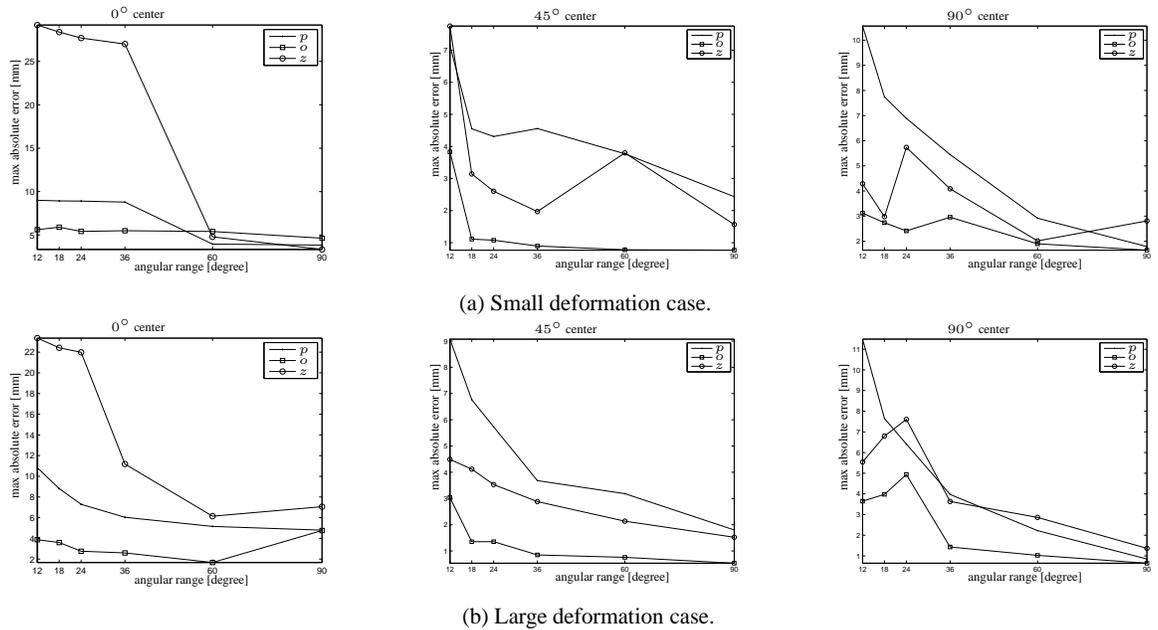


Figure 5.5: Maximum absolute errors (mm) of estimated deformation within ROI. *Left:* Angular center is $\beta = 0^\circ$. *Center:* Angular center is $\beta = 45^\circ$. *Right:* Angular center is $\beta = 90^\circ$.

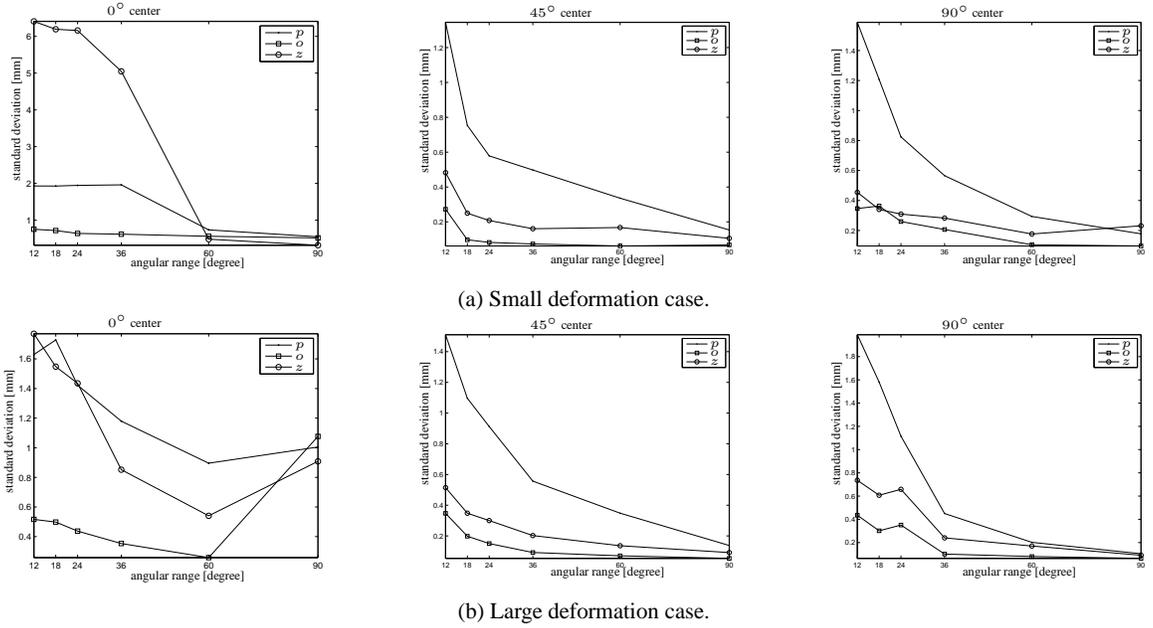


Figure 5.6: Standard deviation of absolute errors (mm) of estimated deformation within ROI. *Left:* Angular center is $\beta = 0^\circ$. *Center:* Angular center is $\beta = 45^\circ$. *Right:* Angular center is $\beta = 90^\circ$.

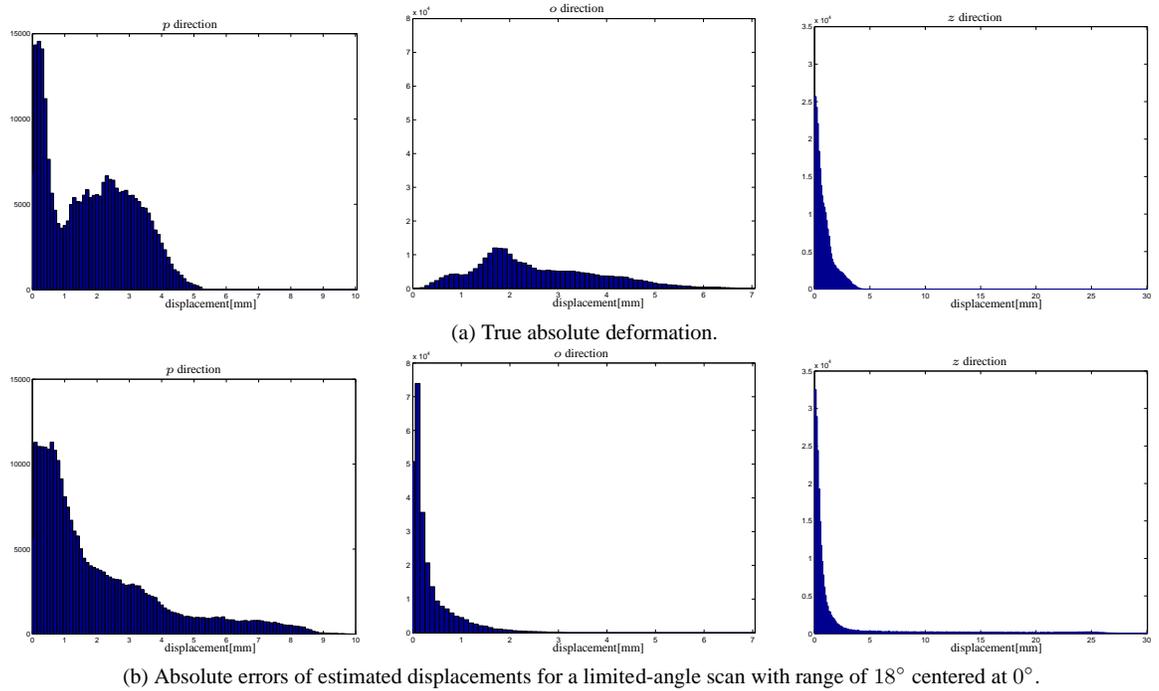


Figure 5.7: Histograms in the o direction (*left*), p direction (*center*) and z direction (*right*) within the ROI for the small deformation case.

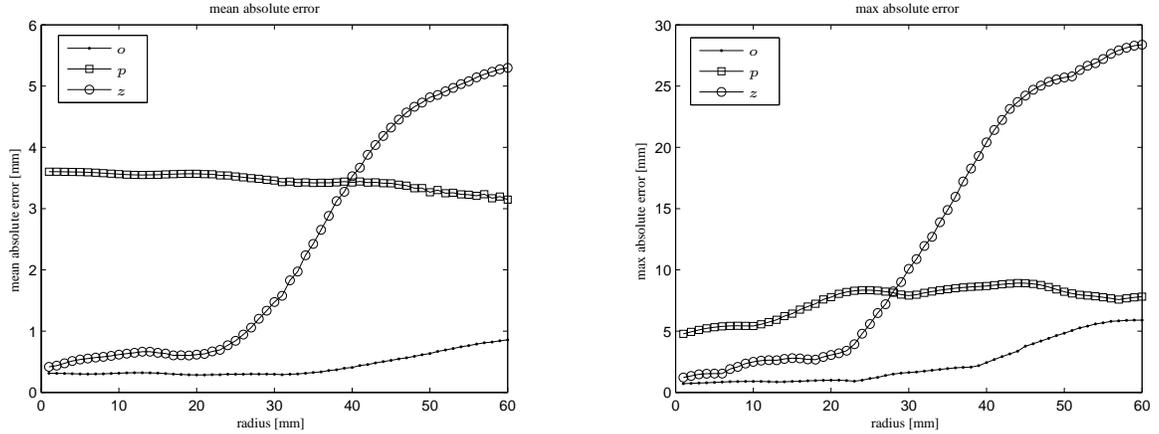


Figure 5.8: Mean (*left*) and maximum (*right*) absolute errors of estimated displacements within ROI for a limited-angle scan with range of 18° centered at 0° . The horizontal axis denotes radial distance to the tumor center. The errors were calculated among points within cylindrical shells centered at the tumor center.

below 0.5mm for angular ranges greater than 36° . For the 0° center cases, the errors and standard deviation are larger than those of 45° and 90° center cases and the error in the z direction is larger than that in the p and o direction.

The maximum absolute error and standard deviation of the limited-angle scans centered at 0° are much larger than those of other cases, especially for the small deformation case with angular range smaller than 36° . We chose the scan with range of 18° centered at 0° for the small deformation case as an example to investigate more details about the estimated deformation.

Fig. 5.7(a) shows histograms of true absolute deformation within the ROI and Fig. 5.7(b) shows histograms of absolute errors of estimated displacements within the ROI for a limited-angle scan with range of 18° centered at 0° for the small deformation case. Only a small percentage of voxels have large absolute errors in the z direction. For this case of a limited-angle scan, we also calculated mean and maximum errors among points within a sequence of cylindrical shells (with thickness of 1mm) centered at the tumor center. The left panel of Fig. 5.1 (d) shows an axial view of 4 such cylinders. Fig. 5.8 shows these errors versus radial distance to the tumor center. With

increasing radial distance, the errors in the z direction increase, especially when radial distance is greater than 30mm. The maximum errors in this direction happened at the bony portion inside the chest wall. The errors in the o direction are small and increase slowly. The errors in the p direction remain at the same level.

The shape information that can be estimated from limited-angle projections depends on the angular center, especially when angular ranges are small. It appears that the truncation of 0° scans affects estimation accuracy more severely than scans at other centers. Since the deformation model covers the whole thorax, truncation outside the ROI still influences estimation accuracy in all regions. This influence becomes more obvious when radial distance from the rotation center (tumor center) increases, *i.e.*, as the distance to truncated parts decreases. One reason for the large errors in z is that the voxel size in z is three times of those in x and y . The poor sampling (not truncation) associated with cone-beam CT geometry at the off-axis slices may also influence accuracy.

5.2.2 Noisy Projection Views

We generated noiseless projection views \mathbf{y}_n of target volumes ($\mathbf{f}_{\text{in}}^{\text{tar}}$ and $\mathbf{f}_{\text{ex}}^{\text{tar}}$) using the distance-driven (DD) method [26], generated transmitted intensities \mathbf{Y}_n using (5.3) with $I_{n,l}$ being 10^5 and 10^4 counts per ray for all n and l , and then estimated noisy projection views $\hat{\mathbf{y}}_n$ using (5.4). The projection-to-volume registration described in Section 5.1.1 was performed between these estimated projections $\hat{\mathbf{y}}_n$ and the reference volume.

We tested estimation accuracy on noisy projections of a limited-angle scan with angular range of 24° centered at 45° of the small deformation case. Table 5.3 shows absolute errors (mm) of tumor center shifts for the tested limited-angle scan. Table 5.4 shows mean, maximum and standard deviation of absolute errors of estimated displacements within ROI for this scan. We denote

	p direction			o direction			z direction		
intensity	I^a	I^b	I^c	I^a	I^b	I^c	I^a	I^b	I^c
error	0.80	0.85	0.91	0.01	0.10	0.46	0.15	0.69	1.45

Table 5.3: Absolute errors (mm) of tumor center shifts for a limited-angle scan with range of 24° centered at 45° of the small deformation case. I^a stands for the case of no noise, I^b stands for the case when $I_{m,n} = 10^5$ counts per ray, and I^c stands for the case when $I_{m,n} = 10^4$ counts per ray.

	p direction			o direction			z direction		
intensity	I^a	I^b	I^c	I^a	I^b	I^c	I^a	I^b	I^c
mean	0.51	0.51	0.64	0.08	0.10	0.18	0.19	0.26	0.41
max	4.31	4.14	4.23	1.08	1.25	1.17	2.60	3.27	3.76
σ	0.58	0.57	0.63	0.08	0.10	0.16	0.21	0.27	0.39

Table 5.4: Mean, maximum and standard deviation σ of absolute errors of estimated displacements within ROI for a limited-angle scan with range of 24° centered at 45° of the small deformation case. I^a stands for the case of no noise, I^b stands for the case when $I_{m,n} = 10^5$ counts per ray, and I^c stands for the case when $I_{m,n} = 10^4$ counts per ray.

noiseless experiments as I^a , noisy experiments with $I_{n,l} = 10^5$ counts per ray as I^b and noisy experiments with $I_{n,l} = 10^4$ counts per ray as I^c in these tables. The errors of experiments I^a are smaller than those of I^b and I^c . The errors increase when incident intensity $I_{n,l}$ decreasing, but the errors in the p direction change very slowly. It appears that the errors in the p direction are dominated by limited shape information that can be extracted from limited-angle projections. The error differences between experiments I^a and I^b are smaller than those between experiments I^b and I^c .

5.3 Conclusion

We also tested a registration method that consists of two steps. The first step was rigid projection-to-volume registration of the whole FOV to roughly align the whole thorax. The second step was projection-to-volume registration of a ROI around the tumor, which only estimates local translation in each direction. However this method did not work well because the rigid FOV registration was not robust to nonrigid motion of the FOV, the translation motion is too simple for movement

within ROI, and the results of second step were very sensitive to the choices of ROIs, such as how big the ROI should be.

The experiments performed demonstrate the significant amount of information present to aid limited angle projection-to-volume alignment. They also highlight some of the trends in degeneracy of such alignments from limited angular samples, both as a function of direction relative to the projection set as well as distance from the rotation center. It is hoped that such experiments can be used to guide optimal development of radiographic alignment and monitoring methods that maximize the prior knowledge available in radiotherapy targeting applications to minimize the time, radiographic dose, and computational resources needed for position monitoring during treatment.

This set of experiments presented both small as well as large scale deformations typically found in the thorax of a radiotherapy patient. While we used a deformation associated with variations in breathing states, we do not propose this methodology as a tool for tracking breathing. It is important to understand the information limits in rotational projection-to-volume registration, as these will impact not only the complexity and operational parameters of positioning or tracking methodologies, but more importantly may indicate optimal design of radiographic localization technology integrated with linear accelerators. As a number of configurations have been proposed and introduced studies such as this may help determine the tradeoffs associated with various parameters such as projection arrangement, noise/dose, and temporal/spatial accuracy limits [11, 90, 109, 140].

CHAPTER 6

Conclusion and Future Work

6.1 Summary

This dissertation addressed issues related to image reconstruction and motion estimation for image-guided radiotherapy (IGRT). We described forward and back projection methods for X-ray cone-beam CT, discussed basis material reconstruction (*e.g.*, soft tissue and bone) using statistical reconstruction methods, and investigated estimation accuracy of limited angle projection-to-volume alignment.

Chapter 3 developed two new approaches for 3D forward and back-projection that we call the separable footprint (SF) projectors: the SF-TR and SF-TT projector. They approximate the voxel footprint functions as 2D separable functions. The separability of these footprint functions greatly simplifies the calculation of their integrals over a detector cell and allows efficient implementation of the SF projectors. The SF-TR projector uses trapezoid functions in the transaxial direction and rectangular functions in the axial direction, whereas the SF-TT projector uses trapezoid functions in both directions. Simulation results showed that the SF-TR projector is more accurate with similar computation speed than the DD projector, and the SF-TT projector is more accurate but computationally slower than the SF-TR projector. The DD projector is particularly favorable relative to other previously published projectors in terms of the balance between speed and accuracy.

To balance accuracy and computation, we recommend to combine the SF-TR and SF-TT method, which is to use the SF-TR projector for voxels corresponding to small cone angles and to use the SF-TT projector for voxels corresponding to larger cone angles.

Chapter 4 proposed a statistical PWLS method with edge-preserving regularizer for two-material decomposition from a single-energy CT scan acquired with differential X-ray filters, split and bow-tie filters. We simulated a split filter with 2 mm aluminum across half the fan beam and 200 μm molybdenum of the other half. The simulated body bow-tie filter contains aluminum, graphite, copper and titanium. We also proposed an optimization transfer method with separable quadratic surrogates to monotonically decrease the PWLS cost function which was non-convex and non-linear. The proposed PWLS method with both filtration reconstructed soft tissue and bone components with 25% and 40% lower RMS errors respectively, and reconstructed density (the sum of soft tissue and bone) images and linear attenuation coefficient images at 511 Kev with 40% lower RMS errors than the JS-FBP method. The PWLS images produced less beam-hardening artifacts and noise than the JS-FBP method. The split and bow-tie filter led to similar results.

Chapter 5 tested the projection-to-volume alignment for estimating the local configuration of a lung tumor, and evaluated the influence of rotational arc length on maximum achievable accuracy. The experiments performed demonstrated the significant amount of information present to aid limited angle projection-to-volume alignment. They also highlighted some of the trends in degeneracy of such alignments from limited angular samples, both as a function of direction relative to the projection set as well as distance from the rotation center. It is hoped that such experiments can be used to guide optimal development of radiographic alignment and monitoring methods that maximize the prior knowledge available in radiotherapy targeting applications to minimize the time, radiographic dose, and computational resources needed for position monitoring during treatment.

6.2 Future Work

- We have implemented the proposed two new SF projectors in ANSI C routine and demonstrated their fast speed. Implementing the SF projectors based on graphics processing unit (GPU) programming techniques could further improve the speed [154]. It needs special considerations of how to perform parallel computing while utilizing the properties of the SF projectors. The SF projectors approximated the voxel footprint functions as 2D separable function that simplified calculation of 2D integral into two 1D integrals that can be implemented simultaneously. Implementing the SF projectors on field-programmable gate array (FPGA) is another future research topic [64, 65].
- We have proposed a statistical PWLS to reconstruct two materials from one single scan using differential filtration. One future work is to extend this method to reconstruct three bases from dual-energy scan acquired with differential filtration. Two component images are not enough for some medical applications liu:09:qio, mendonca:10:mmd, maddah:10:pmv. When quantifying the concentration of iron in a fatty liver, images of three constitute materials, iron, fat and tissue, are required [78]. For the purpose of radiotherapy, in addition to soft-tissue and bone it is also better to know the distribution a third material, such as calcium, metal (gold) or iodine. In principle, two basis materials in a mixture can be accurately separated from DECT measurements [75, 78]. Yu *et al.* [160] provided a third criterion of volume conservation assumption to separate three materials using DECT. The volume conservation assumed that the sum of the volumes of three component materials equals the the volume of their mixture. Considering that volume conservation is not always true, Liu *et al.* [78] assumed mass conservation that the sum of the masses of three materials does not change while mixing them. They proposed a post-reconstruction (image domain) decomposition method,

which first individually reconstructed two images using two sets of measurements at low and high energies, and then obtained basis images from these two reconstructed images. Combining filtration with DECT would create different spectra from the original spectra at two source voltage settings, which would provide additional information for reconstructing three materials.

- The PWLS method proposed in Chapter 4 used edge-preserving regularizers. Each material component has a regularizer coefficient and edge-preserving parameter. We have found the image quality in terms of noise and estimation bias was sensitive to these parameters. The choice of parameters for one material component influenced the reconstructed image of another component. Huh and Fessler [49] used a material-cross penalty for DECT reconstruction. This penalty used the prior knowledge that different component images have similar edge area. Material decomposition could also be considered as a classification problem, since most pixels either have soft tissue characteristics or bone characteristics [29]. Designing penalty functions based on such prior knowledge may improve reconstruction results. Future research could investigate optimal penalty functions and parameters for the PWLS method.
- We have presented an optimization transfer method with separable quadratic surrogates for the PWLS cost with edge-preserving penalty in Chapter 4. For this material decomposition problem, soft tissue and bone values are both estimated for each pixel. It is still practical to have nonseparable quadratic surrogate functions for each pixel because inversion of the 2×2 curvature matrix is doable. The question is now what kind of curvature matrix is optimal. Appendix B described preliminary study on this future work.
- We have simulated an incident spectrum for an X-ray tube voltage at 80 kVp to test the pro-

posed PWLS method for two-material decomposition from one single-energy scan acquired with differential filtration. The spectra produced by our simulated split and bow-tie filter have significant overlap for this 80 kVp voltage setting. However, the spectral differences across rays were sufficient for two-material reconstruction. Future work could investigate the influence of X-ray tube voltages on the degree of spectral separation produced by split and bow-tie X-ray filters and qualities of reconstructed component images. One should consider the radiation dose while choosing different voltages.

- We have showed simulation results of the proposed PWLS method for a 2D fan-beam geometry. In order to apply this method to real data, several practical issues need to be addressed. It requires precisely alignment of the bow-tie or split filter and rotational center in order to expose the isocenter area to different spectra which helps reconstructing component images in that area. Bow-tie filters are commonly used in CT scanners, but using split filters needs further considerations. For 2D fan-beam geometries, each path in the patient cross section is measured with two different beam filtration created by a split filter during a single 360° rotation scan. However, for typical clinical CT systems, such as axial and helical cone beam geometries, most tilted paths are not measured by exactly two different incident spectra.
- Chapter 5 conducted an objective characterization of the influence of rotational arc length on motion estimation accuracy for projection-to-volume targeting during rotational therapy. It was an experimental investigation based on thorax CT volumes of a lung cancer patient. Future research could analyze bounds of estimated displacements from a small spread of projections, such as Bayesian lower bounds [148, 149].
- The experiments performed in Chapter 5 demonstrated the significant amount of information present to aid limited angle projection-to-volume alignment. However, the estimation

errors in the projection direction were bigger than the orthogonal and axial directions since only limited shape information can be extracted from projection views along that direction. Future research could address how to improve the estimation accuracy along the projection direction. Design penalty based on prior knowledge of shape information along the projection direction would improve the estimation accuracy of limited angle projection-to-volume registration. It would also be helpful to use limited-angle projections that acquired along directions that are orthogonal to each other.

APPENDICES

APPENDIX A

Surrogate Function Design

A.1 Local Properties of Beam-Hardening Function

The vector beam-hardening function defined in (4.14) has row gradient $\nabla \mathbf{f}_i$ (a $M_0 \times L_0$ matrix) with elements

$$\frac{\partial}{\partial s_l} f_{im}(\mathbf{s}) = [\nabla \mathbf{f}_i(\mathbf{s})]_{ml} = \frac{-\frac{\partial}{\partial s_l} v_{im}(\mathbf{s})}{v_{im}(\mathbf{s})} = \int q_{im}(\mathcal{E}, \mathbf{s}) \beta_l(\mathcal{E}) d\mathcal{E}, \quad (\text{A.1})$$

where we define the following probability density function:

$$q_{im}(\mathcal{E}; \mathbf{s}) \triangleq p_{im}(\mathcal{E}) e^{-\beta(\mathcal{E}) \cdot \mathbf{s}} / v_{im}(\mathbf{s}).$$

Because the gradient vector $\nabla \mathbf{f}_i$ always has positive elements, f_{im} is a monotone increasing function.

The properties of f_{im} for small material thicknesses (*i.e.*, for $\mathbf{s} \approx \mathbf{0}$) can be useful. In particular:

$$[\nabla \mathbf{f}_i(\mathbf{s})]_{ml} \Big|_{\mathbf{s}=\mathbf{0}} = \bar{\beta}_{iml}, \quad (\text{A.2})$$

where we define the following “effective” mass attenuation coefficient:

$$\bar{\beta}_{iml} \triangleq \int \beta_l(\mathcal{E}) p_{im}(\mathcal{E}) d\mathcal{E}. \quad (\text{A.3})$$

We aggregate these $\bar{\beta}_{iml}$ values into a $M_0 \times L_0$ matrix \mathbf{B}_i having entries

$$[\mathbf{B}_i]_{ml} = \bar{\beta}_{iml}. \quad (\text{A.4})$$

In matrix form, (A.2) becomes

$$\nabla \mathbf{f}_i(\mathbf{s}) \Big|_{\mathbf{s}=\mathbf{0}} = \mathbf{B}_i.$$

In particular, if the source spectra are monoenergetic, *i.e.*, $p_{im}(\mathcal{E}) = \delta(\mathcal{E} - \mathcal{E}_m)$, then $\bar{\beta}_{iml} = \beta_l(\mathcal{E}_m)$ and the \mathbf{f}_i functions are linear in their arguments:

$$\mathbf{f}_i(\mathbf{s}) = \mathbf{B}_i \mathbf{s}, \quad f_{im}(s_i) = \sum_{l=1}^{L_0} \beta_l(\mathcal{E}_m) s_{il}.$$

For polyenergetic measurements, the \mathbf{f}_i functions are somewhat nonlinear, where the degree of nonlinearity depends on the spectrum and the mass attenuation coefficients. In particular, using the concavity of log, one can show that

$$f_{im}(\mathbf{s}) \leq \sum_{l=1}^{L_0} \left(\int p_{im}(\mathcal{E}) \beta_l(\mathcal{E}) d\mathcal{E} \right) s_l = \sum_{l=1}^{L_0} \bar{\beta}_{iml} s_l = [\mathbf{B}_i \mathbf{s}]_m,$$

due to the beam hardening effect. So the f_{im} surface always lies below its tangent plane at zero.

Concavity

One can show that $f_{im}(\mathbf{s})$ is *concave*, *i.e.*, $-\nabla^2 f_{im}(\mathbf{s})$ is a nonnegative definite $L_0 \times L_0$ matrix.

To see this, use (A.1) to show

$$\begin{aligned} -\nabla^2 f_{im}(\mathbf{s}) &= \frac{1}{v_{im}(\mathbf{s})} \nabla^2 v_{im}(\mathbf{s}) - \frac{1}{v_{im}^2(\mathbf{s})} \nabla v_{im}(\mathbf{s}) \nabla v_{im}(\mathbf{s}) \\ &= \frac{1}{v_{im}(\mathbf{s})} \nabla^2 v_{im}(\mathbf{s}) - \nabla f_{im}(\mathbf{s}) \nabla f_{im}(\mathbf{s}) \\ &= \int q_{im}(\mathcal{E}; \mathbf{s}) \beta(\mathcal{E}) \beta'(\mathcal{E}) d\mathcal{E} - \left(\int q_{im}(\mathcal{E}; \mathbf{s}) \beta(\mathcal{E}) d\mathcal{E} \right) \left(\int q_{im}(\mathcal{E}; \mathbf{s}) \beta(\mathcal{E}) d\mathcal{E} \right)'. \end{aligned} \quad (\text{A.5})$$

Because q_{im} is nonnegative and integrates to unity over \mathcal{E} , it is a probability density function. So by Jensen's inequality:

$$\mathbf{x}' [-\nabla^2 f_{im}(\mathbf{s})] \mathbf{x} = \int q_{im}(\mathcal{E}; \mathbf{s}) |\mathbf{x} \cdot \boldsymbol{\beta}(\mathcal{E})|^2 d\mathcal{E} - \left| \int q_{im}(\mathcal{E}; \mathbf{s}) [\mathbf{x} \cdot \boldsymbol{\beta}(\mathcal{E})] d\mathcal{E} \right|^2 \geq 0,$$

where “'” denotes matrix/vector transpose.

Hence the Hessian matrix $-\nabla^2 f_{im}(\mathbf{s})$ is nonnegative definite for any \mathbf{s} , so f_{im} is concave. In addition, usually $\beta_l(\mathcal{E})$ is smallest at \mathcal{E}_{\max} , so the “distribution” $q_{im}(\mathcal{E}; \mathbf{s})$ becomes concentrated around \mathcal{E}_{\max} as \mathbf{s} increases. Thus its “variance,” *i.e.*, $\nabla^2 f_{im}(\mathbf{s})$, approaches zero for large \mathbf{s} , *i.e.*, f_{im} is asymptotically linear [20]. These properties can be useful constraints when fitting approximations to f_{im} .

A.2 Curvature for Quadratic Surrogate

The key to the curvature derivations is the following inequality. When \mathbf{M} is real and symmetric [46, p. 34]: $\mathbf{x}' \mathbf{M} \mathbf{x} \leq \lambda_{\max}(\mathbf{M}) \|\mathbf{x}\|^2$, where $\lambda_{\max}(\mathbf{M})$ denotes the maximum eigenvalue of \mathbf{M} . So we have $\mathbf{M} \preceq \lambda_{\max}(\mathbf{M}) \mathbf{I}$. Because $\nabla f_{im}(\mathbf{s})$ has the largest entries when $\mathbf{s} = \mathbf{0}$:

$$H_1(\mathbf{s}) = \nabla f_{im}(\mathbf{s}) \nabla f_{im}(\mathbf{s}) \preceq \|\nabla f_{im}(\mathbf{s})\|^2 \mathbf{I} \preceq \|\nabla f_{im}(\mathbf{0})\|^2 \mathbf{I} = \|\mathbf{b}_{im}\|^2 \mathbf{I} \triangleq \tilde{\mathbf{C}}_{im},$$

where \mathbf{b}_{im} is the transpose of the m th row of \mathbf{B}_i .

A.3 Separable Quadratic Surrogates Based on Additive Convexity Trick

We form a separable quadratic surrogate function $\phi_2^{(n)}(\mathbf{x})$ with respect to \mathbf{x} by using the following trick due to De Pierro [27]:

$$\mathbf{s}_i(\mathbf{x}) = \sum_{j=1}^{N_p} a_{ij} \mathbf{x}_j = \sum_{j=1}^{N_p} \pi_{ij} \left(\frac{a_{ij}}{\pi_{ij}} (\mathbf{x}_j - \mathbf{x}_j^{(n)}) + \sum_{j=1}^{N_p} a_{ij} \mathbf{x}_j^{(n)} \right), \quad (\text{A.6})$$

provided $\sum_{j=1}^{N_p} \pi_{ij} = 1$ and π_{ij} is zero only if a_{ij} is zero. If the π_{ij} 's are nonnegative, then we can apply the convexity inequality to the quadratic function $q_{im}^{(n)}(\mathbf{s}_i)$ defined in (4.19) to write

$$\begin{aligned} q_{im}^{(n)}(\mathbf{s}_i) &= q_{im}^{(n)}\left(\sum_{j=1}^{N_p} a_{ij} \mathbf{x}_j\right) \\ &= q_{im}^{(n)}\left(\sum_{j=1}^{N_p} \pi_{ij} \left(\frac{a_{ij}}{\pi_{ij}}(\mathbf{x}_j - \mathbf{x}_j^{(n)}) + \sum_{j=1}^{N_p} a_{ij} \mathbf{x}_j^{(n)}\right)\right) \\ &\leq \sum_{j=1}^{N_p} \pi_{ij} q_{im}^{(n)}\left(\frac{a_{ij}}{\pi_{ij}}(\mathbf{x}_j - \mathbf{x}_j^{(n)}) + \sum_{j=1}^{N_p} a_{ij} \mathbf{x}_j^{(n)}\right). \end{aligned} \quad (\text{A.7})$$

Combining these yields the following separable quadratic surrogate for $\Psi(\mathbf{x})$:

$$\phi_1^{(n)}(\mathbf{s}_i) \leq \phi_2^{(n)}(\mathbf{x}) \triangleq \sum_{j=1}^{N_p} \phi_j^{(n)}(\mathbf{x}_j), \quad (\text{A.8})$$

where

$$\phi_j^{(n)}(\mathbf{x}_j) = \sum_{i=1}^{N_d} \sum_{m=1}^{M_0} \pi_{ij} q_{im}^{(n)}\left(\frac{a_{ij}}{\pi_{ij}}(\mathbf{x}_j - \mathbf{x}_j^{(n)}) + \sum_{j=1}^{N_p} a_{ij} \mathbf{x}_j^{(n)}\right). \quad (\text{A.9})$$

The column gradient $\nabla \phi_j^{(n)}$ has elements

$$\frac{\partial}{\partial x_{lj}} \phi_j^{(n)}(\mathbf{x}_j) = \sum_{i=1}^{N_d} \sum_{m=1}^{M_0} a_{ij} \dot{q}_{im}^{(n)}\left(\frac{a_{ij}}{\pi_{ij}}(\mathbf{x}_j - \mathbf{x}_j^{(n)}) + \sum_{j=1}^{N_p} a_{ij} \mathbf{x}_j^{(n)}\right), \quad (\text{A.10})$$

so

$$\begin{aligned} \left. \frac{\partial}{\partial x_{lj}} \phi_j^{(n)}(\mathbf{x}_j) \right|_{\mathbf{x}_j = \mathbf{x}_j^{(n)}} &= \sum_{i=1}^{N_d} \sum_{m=1}^{M_0} a_{ij} \dot{q}_{im}^{(n)}\left(\sum_{j=1}^{N_p} a_{ij} \mathbf{x}_j^{(n)}\right) \\ &= \sum_{i=1}^{N_d} \sum_{m=1}^{M_0} a_{ij} \dot{\Psi}_{im}\left(\sum_{j=1}^{N_p} a_{ij} \mathbf{x}_j^{(n)}\right) \\ &= \frac{\partial}{\partial x_{lj}} \Psi(\mathbf{x}^{(n)}). \end{aligned} \quad (\text{A.11})$$

The hessian $D_j^{(n)} \triangleq \nabla^2 \phi_j^{(n)}$ has elements

$$\begin{aligned}
\left[D_j^{(n)} \right]_{lk} &= \frac{\partial^2}{\partial x_{lj} \partial x_{kj}} \phi_j^{(n)}(\mathbf{x}_j) \\
&= \sum_{i=1}^{N_d} \sum_{m=1}^{M_0} \frac{a_{ij}^2}{\pi_{ij}} \frac{\partial^2}{\partial x_{lj} \partial x_{kj}} q_{im}^{(n)} \left(\frac{a_{ij}}{\pi_{ij}} (\mathbf{x}_j - \mathbf{x}_j^{(n)}) + \sum_{j=1}^{N_p} a_{ij} \mathbf{x}_j^{(n)} \right) \\
&= \sum_{i=1}^{N_d} \sum_{m=1}^{M_0} \frac{a_{ij}^2}{\pi_{ij}} \left[\tilde{C}_{im} \right]_{lk}, \tag{A.12}
\end{aligned}$$

so

$$D_j^{(n)} = \sum_{i=1}^{N_d} \sum_{m=1}^{M_0} \frac{a_{ij}^2}{\pi_{ij}} \tilde{C}_{im}. \tag{A.13}$$

Combining the gradient and hessian, the separable quadratic surrogate satisfies the conditions (2.67) has the form

$$\phi_2^{(n)}(\mathbf{x}) = \Psi(\mathbf{x}^{(n)}) + \nabla \Psi(\mathbf{x}^{(n)}) (\mathbf{x} - \mathbf{x}^{(n)}) + \frac{1}{2} (\mathbf{x} - \mathbf{x}^{(n)})' \text{diag} \left\{ D_j^{(n)} \right\} (\mathbf{x} - \mathbf{x}^{(n)}), \tag{A.14}$$

where $\text{diag} \left\{ D_j^{(n)} \right\}$ is a block diagonal matrix. Because this $\phi_2^{(n)}(\mathbf{x})$ is separable quadratic function, it is trivial to minimize, leading to the following parallelizable iteration:

$$\mathbf{x}_j^{(n+1)} = \left[\mathbf{x}_j^{(n)} - \left[D_j^{(n)} \right]^{-1} \nabla_{\mathbf{x}_j} \Psi(\mathbf{x}^{(n)}) \right]_+, \quad j = 1, \dots, N_p, \tag{A.15}$$

where $\nabla_{\mathbf{x}_j}$ denotes the gradient with respect to \mathbf{x}_j . In matrix-vector form, the update is:

$$\mathbf{x}^{(n+1)} = \left[\mathbf{x}^{(n)} - \text{diag} \left\{ \left[D_j^{(n)} \right]^{-1} \right\} \nabla \Psi(\mathbf{x}^{(n)}) \right]_+, \tag{A.16}$$

which is a kind of diagonally-preconditioned gradient descent algorithm that is guaranteed to monotonically decrease the cost function each iteration. This algorithm is entirely parallelizable because all pixels can be updated simultaneously.

A.3.1 Curvature of Diagonal Matrix

When the curvature $\tilde{\mathbf{C}}_{im}$ given in (4.24) is a diagonal matrix, the separable surrogate and update described in Section A.3 can be further simplified. Define the l th element of $D_j^{(n)}$ as

$$d_{lj}^{(n)} \triangleq [D_j^{(n)}]_{ll} = \sum_{i=1}^{N_d} \sum_{m=1}^{M_0} \frac{a_{ij}^2}{\pi_{ij}} [\tilde{\mathbf{C}}_{im}]_{ll}. \quad (\text{A.17})$$

The separable quadratic surrogate in (A.14) becomes

$$\phi_2^{(n)}(\mathbf{x}) = \Psi(\mathbf{x}^{(n)}) + \nabla \Psi(\mathbf{x}^{(n)}) (\mathbf{x} - \mathbf{x}^{(n)}) + \frac{1}{2} (\mathbf{x} - \mathbf{x}^{(n)})' \text{diag}\{d_{lj}^{(n)}\} (\mathbf{x} - \mathbf{x}^{(n)}), \quad (\text{A.18})$$

and the update in (A.16) becomes

$$\mathbf{x}^{(n+1)} = \left[\mathbf{x}^{(n)} - \text{diag}\left\{ \frac{1}{d_{lj}^{(n)}} \right\} \nabla \Psi(\mathbf{x}^{(n)}) \right]_+. \quad (\text{A.19})$$

APPENDIX B

Optimal Curvature for Two-Dimensional Quadratic Surrogates

Optimization transfer methods convert the problem of optimizing difficult cost functions to optimizing a sequence of relatively simpler surrogate functions. They take into account specific properties of cost functions of interest, and are guaranteed to monotonically decrease the cost function if suitable surrogate functions are designed. Quadratic surrogates are often desired because there is a simple closed form solution at the “M-step” (2.68) in the absence of constraints. To achieve fast convergence rate, one wants curvatures of quadratic surrogate functions to be as small as possible while satisfying the surrogate conditions (2.67). Since it is usually difficult to find surrogate functions that fit all the desired conditions, the choice of surrogate functions is something of an art.

Due to the huge size of imaging problems, separable quadratic surrogates are particularly appealing because one can update all pixels simultaneously. For separable surrogates, the design problem simplifies to finding a suitable 1D surrogate function. Let $\psi(t)$ denote the cost function and $h(t; s)$ denote a surrogate function that satisfies the following two conditions:

$$\begin{aligned} h(s; s) &= \psi(s), & \forall s \\ h(t; s) &\geq \psi(t), & \forall t, s. \end{aligned} \tag{B.1}$$

For a 1D quadratic surrogate function h , the optimal curvature can be defined to be the smallest curvature that still ensures the surrogate conditions (B.1):

$$\check{c}_{\text{opt}}(\psi, s) \triangleq \min \left\{ c \geq 0 : \psi(s) + \dot{\psi}(s)(t-s) + \frac{1}{2}c(t-s)^2 \geq \psi(t), \forall t \right\}. \quad (\text{B.2})$$

For some applications, several related images are estimated at the same time. For example, DECT methods estimate soft tissue and bone values for each pixel. It can still be practical to have nonseparable quadratic surrogate functions for each pixel because inversion of a 2×2 or 3×3 curvature matrix is doable. The question is now what kind of curvature matrix is optimal. In this appendix, we define the optimal curvature matrices, discuss the surrogate conditions and optimal criteria, and derive optimal curvature matrices for several special cases.

B.1 Definition of An Optimal Curvature Matrix

Let $\phi^{(n)}$ be a quadratic surrogate function for a cost function $\Psi(\mathbf{x})$ at the n th iteration, having the following form:

$$\phi^{(n)}(\mathbf{x}) = \Psi(\mathbf{x}^{(n)}) + (\mathbf{x} - \mathbf{x}^{(n)})' \nabla \Psi(\mathbf{x}^{(n)}) + \frac{1}{2}(\mathbf{x} - \mathbf{x}^{(n)})' \mathbf{C}_n (\mathbf{x} - \mathbf{x}^{(n)}). \quad (\text{B.3})$$

We define the optimal curvature matrix $\tilde{\mathbf{C}}_{\text{opt}}(\Psi, \mathbf{x}^{(n)}; \mathcal{C})$ for a quadratic surrogate $\phi^{(n)}$ as follows:

1. Pick a structure \mathcal{C} for \mathbf{C}_n , *e.g.*, separable quadratic surrogates

$$\mathcal{C} = \left\{ \mathbf{C}_n = \begin{bmatrix} \alpha_1 & 0 \\ 0 & \alpha_2 \end{bmatrix} : \alpha_1, \alpha_2 \geq 0 \right\}. \quad (\text{B.4})$$

2. Require the quadratic surrogate $\phi^{(n)}$ to be a majorizer of Ψ , *i.e.*

$$\mathcal{C}_m = \{ \mathbf{C}_n \in \mathcal{C} : \phi^{(n)}(\mathbf{x}) \geq \Psi(\mathbf{x}), \forall \mathbf{x} \}. \quad (\text{B.5})$$

Due to their form, quadratic surrogate functions (B.3) always satisfy the equality surrogate condition.

3. Pick the optimal curvature by some criterion, *e.g.*

$$\tilde{\mathbf{C}}_{\text{opt}}(\Psi, \mathbf{x}^{(n)}; \mathcal{C}) \triangleq \arg \min_{\mathbf{C}_n \in \mathcal{C}_m} \rho(\mathbf{I} - \mathbf{C}_n^{-1} \nabla^2 \Psi(\mathbf{x}^{(n)})), \quad (\text{B.6})$$

where ρ denotes the spectral radius. The criterion (B.6) is a natural choice because it is related to convergence rate.

DECT methods are our interest, so we focus on 2D surrogate functions hereafter, *i.e.*, $\mathbf{x} \in \mathbb{R}^2$.

B.2 Majorization Constraint

For a twice differentiable function $\Psi(\mathbf{x})$, its 2nd-order Taylor series expansion about the current estimate $\mathbf{x}^{(n)}$ is

$$\begin{aligned} \Psi(\mathbf{x}) &= \Psi(\mathbf{x}^{(n)}) + (\mathbf{x} - \mathbf{x}^{(n)})' \nabla \Psi(\mathbf{x}^{(n)}) \\ &\quad + (\mathbf{x} - \mathbf{x}^{(n)})' \left[\int_0^1 (1 - \alpha) \nabla^2 \Psi(\alpha \mathbf{x} + (1 - \alpha) \mathbf{x}^{(n)}) d\alpha \right] (\mathbf{x} - \mathbf{x}^{(n)}). \end{aligned} \quad (\text{B.7})$$

For $\phi^{(n)}$ in (B.3) to be a valid quadratic surrogate, it is sufficient that \mathbf{C}_n satisfies the following conditions:

$$\mathbf{C}_n \succeq \nabla^2 \Psi(\mathbf{x}), \quad \forall \mathbf{x} \quad (\text{B.8})$$

$$\mathbf{C}_n \succeq \mathbf{0}. \quad (\text{B.9})$$

Let \mathcal{C}_s denote the set of matrices satisfying the above sufficient conditions:

$$\mathcal{C}_s = \{ \mathbf{C}_n \succeq \mathbf{0} : \mathbf{C}_n \succeq \nabla^2 \Psi(\mathbf{x}), \forall \mathbf{x} \}. \quad (\text{B.10})$$

We hereafter focus on conditions (B.8) and (B.9) ($\mathcal{C}_s \subseteq \mathcal{C}_m$), because they readily ensure the majorization constraint given in (B.5).

Typically the curvature matrix C_n is designed to be a symmetric positive semidefinite matrix, *i.e.*, satisfying constraint (B.9). Let z_1 and z_2 be the orthonormal eigenvectors of C_n , and $\alpha = (\alpha_1, \alpha_2)$ where α_1 and α_2 are the corresponding eigenvalues. Let $Q = \begin{bmatrix} z_1 & z_2 \end{bmatrix}$, then $QQ' = Q'Q = I$. We can write C_n as

$$C_n(\alpha, Q) = \alpha_1 z_1 z_1' + \alpha_2 z_2 z_2' = Q\Lambda(\alpha)Q', \quad (\text{B.11})$$

where $\Lambda(\alpha) = \begin{bmatrix} \alpha_1 & 0 \\ 0 & \alpha_2 \end{bmatrix}$. The condition (B.8) becomes

$$\begin{aligned} C_n(\alpha, Q) &\succeq \nabla^2 \Psi(x), \quad \forall x \\ \iff \mathbf{y}'Q\Lambda(\alpha)Q'\mathbf{y} &\geq \mathbf{y}'\nabla^2 \Psi(x)\mathbf{y}, \quad \forall \mathbf{y}, \forall x \\ \iff \mathbf{u}'\Lambda(\alpha)\mathbf{u} &\geq \mathbf{u}'Q'\nabla^2 \Psi(x)Q\mathbf{u}, \quad \text{where } \mathbf{u} = Q'\mathbf{y}, \forall \mathbf{u}, x \\ \iff \mathbf{u}'T(\alpha, x)\mathbf{u} &\geq 0, \quad \forall \mathbf{u}, \forall x \\ \iff \lambda_{\min}(T(\alpha, x)) &\geq 0, \quad \forall x, \end{aligned} \quad (\text{B.12})$$

where

$$T(\alpha, x) \triangleq \Lambda(\alpha) - Q'\nabla^2 \Psi(x)Q. \quad (\text{B.13})$$

Here, λ denotes an eigenvalue of $T(\alpha, x)$ and λ_{\min} denotes the smallest one, *i.e.*, $\lambda = \text{eig}\{T(\alpha, x)\}$ and $\lambda_{\min} = \lambda_{\min}(T(\alpha, x)) = \min\{\text{eig}\{T(\alpha, x)\}\}$. The last equivalence in (B.12) follows from the fact that $T(\alpha, x)$ is hermitian symmetric.

B.2.1 Diagonal curvature matrix

The general form of the hessian is $\nabla^2 \Psi(x) = \begin{bmatrix} a(x) & b(x) \\ b(x) & c(x) \end{bmatrix}$, where $a(x), b(x), c(x) \in \mathbb{R}$.

To simplify the problem, we consider separable surrogate functions, *i.e.*, $Q = I$. Their curvature

matrices are diagonal matrices, *i.e.*, $C_n(\boldsymbol{\alpha}, \mathbf{Q}) = \Lambda(\boldsymbol{\alpha})$. The matrix defined in (B.13) becomes

$$\mathbf{T}(\boldsymbol{\alpha}, \mathbf{x}) = \Lambda(\boldsymbol{\alpha}) - \nabla^2 \Psi(\mathbf{x}) = \begin{bmatrix} \alpha_1 - a(\mathbf{x}) & -b(\mathbf{x}) \\ -b(\mathbf{x}) & \alpha_2 - c(\mathbf{x}) \end{bmatrix}.$$

The determinant of $\mathbf{T}(\boldsymbol{\alpha}, \mathbf{x}) - \lambda \mathbf{I}$ is

$$(\alpha_1 - a(\mathbf{x}) - \lambda)(\alpha_2 - c(\mathbf{x}) - \lambda) - b^2(\mathbf{x}) = \lambda^2 + B\lambda + C,$$

where

$$\begin{aligned} B &= B(\boldsymbol{\alpha}, \mathbf{x}) \triangleq a(\mathbf{x}) + c(\mathbf{x}) - \alpha_1 - \alpha_2 \\ C &= C(\boldsymbol{\alpha}, \mathbf{x}) \triangleq (\alpha_1 - a(\mathbf{x}))(\alpha_2 - c(\mathbf{x})) - b(\mathbf{x})^2. \end{aligned}$$

Let $\det\{\mathbf{T}(\boldsymbol{\alpha}, \mathbf{x}) - \lambda \mathbf{I}\} = 0$, and then

$$\lambda = \frac{1}{2} \left(-B \pm \sqrt{B^2 - 4C} \right) \Rightarrow \lambda_{\min} = \frac{1}{2} \left(-B - \sqrt{B^2 - 4C} \right),$$

where

$$B^2 - 4C = (-\alpha_1 + \alpha_2 + a(\mathbf{x}) - c(\mathbf{x}))^2 + 4b^2(\mathbf{x}) \geq 0.$$

Working on condition (B.12), we have

$$\begin{aligned} &\lambda_{\min} \geq 0 \\ \iff &-B - \sqrt{B^2 - 4C} \geq 0 \\ \iff &\sqrt{B^2 - 4C} \leq -B, \quad \text{so } B \leq 0 \\ \iff &B^2 - 4C \leq B^2 \quad \text{and} \quad B \leq 0 \\ \iff &C \geq 0 \quad \text{and} \quad B \leq 0. \end{aligned} \tag{B.14}$$

Thus for general Hessians, diagonal curvature matrix $C_n(\boldsymbol{\alpha}, \mathbf{Q}) = \Lambda(\boldsymbol{\alpha})$ is in \mathcal{C}_s if and only if

$$(\alpha_1 - a(\mathbf{x}))(\alpha_2 - c(\mathbf{x})) \geq b^2(\mathbf{x}), \quad \forall \mathbf{x} \quad (\text{B.15})$$

$$\alpha_1 + \alpha_2 \geq a(\mathbf{x}) + c(\mathbf{x}) \quad \forall \mathbf{x}. \quad (\text{B.16})$$

We use these conditions to find appropriate α_1 and α_2 values.

B.2.2 Curvature of Constant Times the Identity Matrix

In this section, we discuss a special case of diagonal curvature matrix: matrix of constant times the identity matrix. Let \mathbf{C}_n have the following structure, *i.e.*,

$$\mathbf{C}_n \in \mathcal{C}, \quad \mathcal{C} = \{\mathbf{C}_n = \alpha \mathbf{I} : \alpha \geq 0\}. \quad (\text{B.17})$$

The majorization conditions (B.15) and (B.16) simplify to

$$\alpha^2 - \alpha(a(\mathbf{x}) + c(\mathbf{x})) + a(\mathbf{x})c(\mathbf{x}) - b^2(\mathbf{x}) \geq 0, \quad \forall \mathbf{x} \quad (\text{B.18})$$

$$\frac{1}{2}(a(\mathbf{x}) + c(\mathbf{x})) \leq \alpha, \quad \forall \mathbf{x}. \quad (\text{B.19})$$

Simplifying (B.18) further leads to

$$\begin{aligned} \alpha &\geq \frac{1}{2}(a(\mathbf{x}) + c(\mathbf{x})) + \frac{1}{2}\sqrt{(a(\mathbf{x}) - c(\mathbf{x}))^2 + 4b^2(\mathbf{x})}, \quad \forall \mathbf{x} \\ \text{or} \quad \alpha &\leq \frac{1}{2}(a(\mathbf{x}) + c(\mathbf{x})) - \frac{1}{2}\sqrt{(a(\mathbf{x}) - c(\mathbf{x}))^2 + 4b^2(\mathbf{x})}, \quad \forall \mathbf{x}. \end{aligned} \quad (\text{B.20})$$

Combining (B.20) and (B.19), we have

$$\alpha \geq \frac{1}{2}(a(\mathbf{x}) + c(\mathbf{x})) + \frac{1}{2}\sqrt{(a(\mathbf{x}) - c(\mathbf{x}))^2 + 4b^2(\mathbf{x})}, \quad \forall \mathbf{x}. \quad (\text{B.21})$$

Intuitively, for fast convergence we want α to be as small as possible subject to (B.21). This is shown rigorously in following sections.

B.3 Minimum Spectral Radius Criterion

The local convergence rate of the quadratic surrogate (B.3) is governed by the spectral radius of $\mathbf{I} - \mathbf{C}_n^{-1} \nabla^2 \Psi(\mathbf{x}^{(n)})$. It involves inverting the curvature matrix. For simplicity we focus here on the minimum spectral radius for separable quadratic surrogates, *i.e.*, $\mathbf{C}_n = \begin{bmatrix} \alpha_1 & 0 \\ 0 & \alpha_2 \end{bmatrix}$ where $\alpha_1, \alpha_2 \geq 0$. Let the hessian matrix have the general form $\nabla^2 \Psi(\mathbf{x}^{(n)}) = \begin{bmatrix} a_n & b_n \\ b_n & c_n \end{bmatrix}$ where $a_n = a(\mathbf{x}^{(n)})$, $b_n = b(\mathbf{x}^{(n)})$, $c_n = c(\mathbf{x}^{(n)})$. In this section, we ignore the majorization constraint (B.5), and solve the following minimization problem:

$$(\hat{\alpha}_1, \hat{\alpha}_2) = \arg \min_{\alpha_1, \alpha_2 \geq 0} \rho(\mathbf{D}), \quad \mathbf{D} \triangleq \mathbf{D}(\alpha_1, \alpha_2) = \mathbf{I} - \begin{bmatrix} \alpha_1 & 0 \\ 0 & \alpha_2 \end{bmatrix}^{-1} \begin{bmatrix} a_n & b_n \\ b_n & c_n \end{bmatrix}. \quad (\text{B.22})$$

The minimum value $\rho(\mathbf{D}(\alpha_1, \alpha_2))$ provides a lower bound on the root convergence factor.

B.3.1 Diagonal Curvature Matrix

Let λ be an eigenvalue of \mathbf{D} , then $\rho(\mathbf{D}) = \max_{\lambda \in \text{eig}\{\mathbf{D}\}} |\lambda|$. The determinant of $(\mathbf{D} - \lambda \mathbf{I})$ is

$$\begin{aligned} \det\{\mathbf{D} - \lambda \mathbf{I}\} &= \det \left\{ \begin{bmatrix} 1 - \frac{a_n}{\alpha_1} - \lambda & -\frac{b_n}{\alpha_1} \\ -\frac{b_n}{\alpha_2} & 1 - \frac{c_n}{\alpha_2} - \lambda \end{bmatrix} \right\} \\ &= \left(1 - \frac{a_n}{\alpha_1} - \lambda\right) \left(1 - \frac{c_n}{\alpha_2} - \lambda\right) - \frac{b_n^2}{\alpha_1 \alpha_2} \\ &= \lambda^2 + B_1 \lambda + C_1, \end{aligned} \quad (\text{B.23})$$

where

$$\begin{aligned} B_1 &\triangleq B_1(\alpha_1, \alpha_2) = \frac{a_n}{\alpha_1} + \frac{c_n}{\alpha_2} - 2 \\ C_1 &\triangleq C_1(\alpha_1, \alpha_2) = \left(1 - \frac{a_n}{\alpha_1}\right) \left(1 - \frac{c_n}{\alpha_2}\right) - \frac{b_n^2}{\alpha_1 \alpha_2}. \end{aligned} \quad (\text{B.24})$$

The eigenvalues satisfy $\det\{\mathbf{D} - \lambda\mathbf{I}\} = 0$, and are

$$\lambda = \frac{1}{2} \left(-B_1 \pm \sqrt{B_1^2 - 4C_1} \right) \Rightarrow \rho = \frac{1}{2} \left(|B_1| + \sqrt{B_1^2 - 4C_1} \right),$$

where

$$B_1^2 - 4C_1 = \left[\left(1 - \frac{a_n}{\alpha_1} \right) - \left(1 - \frac{c_n}{\alpha_2} \right) \right]^2 + \frac{4b_n^2}{\alpha_1\alpha_2} \geq 0.$$

The minimization problem becomes

$$\left(\hat{\beta}_1, \hat{\beta}_2 \right) = \arg \min_{\beta_1, \beta_2 \in \mathbb{R}} f(\beta_1, \beta_2), \quad f(\beta_1, \beta_2) = |2 - \beta_1 - \beta_2| + \sqrt{(\beta_1 - \beta_2)^2 + 4d_n\beta_1\beta_2}, \quad (\text{B.25})$$

where

$$\beta_1 = \frac{a_n}{\alpha_1}, \quad \beta_2 = \frac{c_n}{\alpha_2}, \quad d_n = \frac{b_n^2}{a_n c_n}.$$

Let $f_1(\beta_1, \beta_2) = |2 - \beta_1 - \beta_2|$ and $f_2(\beta_1, \beta_2) = \sqrt{(\beta_1 - \beta_2)^2 + 4d_n\beta_1\beta_2}$.

$f_2(\beta_1, \beta_2)$ is symmetric with respect to β_1 and β_2 . The first order partial derivative of $f_2(\beta_1, \beta_2)$

with respect to β_1 is

$$\frac{\partial}{\partial \beta_1} f_2(\beta_1, \beta_2) = \frac{(\beta_1 - \beta_2) + 2d_n\beta_2}{\sqrt{(\beta_1 - \beta_2)^2 + 4d_n\beta_1\beta_2}}.$$

The second order partial derivative of $f_2(\beta_1, \beta_2)$ with respect to β_1 is

$$\begin{aligned} \frac{\partial^2}{\partial \beta_1^2} f_2(\beta_1, \beta_2) &= \frac{\sqrt{(\beta_1 - \beta_2)^2 + 4d_n\beta_1\beta_2} - [(\beta_1 - \beta_2) + 2d_n\beta_2] \frac{(\beta_1 - \beta_2) + 2d_n\beta_2}{\sqrt{(\beta_1 - \beta_2)^2 + 4d_n\beta_1\beta_2}}}{(\beta_1 - \beta_2)^2 + 4d_n\beta_1\beta_2} \\ &= \frac{4d_n\beta_2^2(1 - d_n)}{[(\beta_1 - \beta_2)^2 + 4d_n\beta_1\beta_2]^{\frac{3}{2}}}. \end{aligned}$$

Assume $\mathbf{x}^{(n)}$ is a local minimum of $\Psi(\mathbf{x})$ where Ψ is locally strictly convex, then $a_n c_n > b_n^2$ and $a_n, c_n > 0$, so $0 \leq d_n < 1$. Thus $f_2(\beta_1, \beta_2)$ is strictly convex and $f(\beta_1, \beta_2)$ is strictly convex because $f_1(\beta_1, \beta_2)$ is also strictly convex. It has a unique minimum which has to be at $\beta_1 = \beta_2$

because it is symmetric with respect to β_1 and β_2 . Let $z = \beta_1 = \beta_2$, then rewrite $f(\beta_1, \beta_2)$ as

$$\begin{aligned} f(z) &= 2|1-z| + 2|z|\sqrt{d_n} \\ &= \begin{cases} 2z(\sqrt{d_n}+1) - 2, & z \geq 1 \\ 2z(\sqrt{d_n}-1) + 2, & 0 \leq z \leq 1 \\ -2z(\sqrt{d_n}+1) + 2, & z \leq 0 \end{cases} \end{aligned} \quad (\text{B.26})$$

Fig. B.1 shows this 1D function $f(z)$ when $d_n = 0.5$. More generally, it has a unique minimum at $z = 1$ for all $0 \leq d_n < 1$. Therefore, the solution to (B.22) is

$$\hat{\alpha}_1 = a_n, \quad \hat{\alpha}_2 = c_n. \quad (\text{B.27})$$

The root convergence factor for this optimum is

$$\rho_2 = \frac{1}{2}f(1) = \sqrt{d_n}. \quad (\text{B.28})$$

Unfortunately, in general this design does not necessarily satisfy $\mathbf{C}_n \succeq \nabla^2 \Psi(\mathbf{x}), \forall \mathbf{x}$, but it does majorize $\Psi(\mathbf{x})$ locally to $\mathbf{x}^{(n)}$.

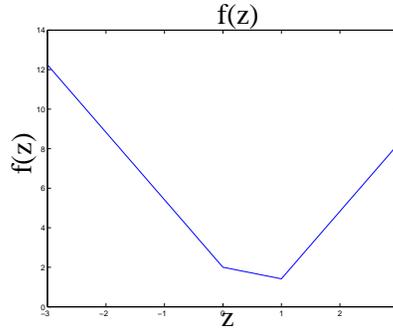


Figure B.1: $f(z)$ when $d_n = 0.5$.

B.3.2 Curvature of Constant Times the Identity Matrix

In this section, we discuss a special case of diagonal curvature matrix: matrix of constant times the identity matrix. Let \mathbf{C}_n have the following structure, *i.e.*,

$$\mathbf{C}_n \in \mathcal{C}, \quad \mathcal{C} = \{\mathbf{C}_n = \alpha \mathbf{I} : \alpha \geq 0\} \quad (\text{B.29})$$

The minimizing problem (B.22) simplifies to

$$\hat{\alpha} = \arg \min_{\alpha \geq 0} \rho(\mathbf{D}(\alpha)), \quad \mathbf{D} \triangleq \mathbf{D}(\alpha) = \mathbf{I} - \frac{1}{\alpha} \begin{bmatrix} a_n & b_n \\ b_n & c_n \end{bmatrix}. \quad (\text{B.30})$$

Let λ be an eigenvalue of \mathbf{D} , then $\rho(\mathbf{D}) = \max_{\lambda \in \text{eig}\{\mathbf{D}\}} |\lambda|$. The determinant of $(\mathbf{D} - \lambda \mathbf{I})$ is

$$\begin{aligned} \det\{\mathbf{D} - \lambda \mathbf{I}\} &= \det \left\{ \begin{bmatrix} 1 - \frac{a_n}{\alpha} - \lambda & -\frac{b_n}{\alpha} \\ -\frac{b_n}{\alpha} & 1 - \frac{c_n}{\alpha} - \lambda \end{bmatrix} \right\} \\ &= \left(1 - \frac{a_n}{\alpha} - \lambda\right) \left(1 - \frac{c_n}{\alpha} - \lambda\right) - \frac{b_n^2}{\alpha^2} \\ &= \lambda^2 + B_2 \lambda + C_2, \end{aligned}$$

where

$$\begin{aligned} B_2 &\triangleq B_2(\alpha) = \frac{a_n + c_n}{\alpha} - 2 \\ C_2 &\triangleq C_2(\alpha) = \left(1 - \frac{a_n}{\alpha}\right) \left(1 - \frac{c_n}{\alpha}\right) - \frac{b_n^2}{\alpha^2}. \end{aligned}$$

Let $\det\{\mathbf{D} - \lambda \mathbf{I}\} = 0$, and then

$$\lambda = \frac{1}{2} \left(-B_2 \pm \sqrt{B_2^2 - 4C_2} \right) \Rightarrow \rho = \frac{1}{2} \left(|B_2| + \sqrt{B_2^2 - 4C_2} \right),$$

where

$$B_2^2 - 4C_2 = \left[\left(1 - \frac{a_n}{\alpha}\right) - \left(1 - \frac{c_n}{\alpha}\right) \right]^2 + \frac{4b_n^2}{\alpha^2} \geq 0.$$

The minimizing problem (B.30) becomes

$$\hat{\alpha} = \arg \min_{\alpha \geq 0} f(\alpha), \quad f(\alpha) = \left| 2 - \frac{a_n + c_n}{\alpha} \right| + \frac{\sqrt{(a_n - c_n)^2 + 4b_n^2}}{\alpha}. \quad (\text{B.31})$$

We solve it as follows.

- When $(2 - \frac{a_n+c_n}{\alpha}) \geq 0 \Rightarrow \alpha \geq \frac{1}{2}(a_n + c_n)$

$$\begin{aligned} f(\alpha) &= 2 + \frac{\sqrt{(a_n - c_n)^2 + 4b_n^2} - (a_n + c_n)}{\alpha} \\ &= 2 + \frac{\sqrt{(a_n + c_n)^2 + 4(b_n^2 - ac)} - (a_n + c_n)}{\alpha}. \end{aligned} \quad (\text{B.32})$$

Assume $\mathbf{x}^{(n)}$ is a local minimum of $\Psi(\mathbf{x})$ where Ψ is locally strictly convex, then

$$ac > b_n^2, a_n > 0, c_n > 0 \Rightarrow \sqrt{(a_n + c_n)^2 + 4(b_n^2 - ac)} - (a_n + c_n) < 0.$$

$f(\alpha)$ is an increasing function of α . Therefore, $\arg \min_{\alpha} f(\alpha) = \frac{1}{2}(a_n + c_n)$ when $\alpha \geq \frac{1}{2}(a_n + c_n)$.

- When $(2 - \frac{a_n+c_n}{\alpha}) \leq 0 \Rightarrow \alpha \leq \frac{1}{2}(a_n + c_n)$

$$f(\alpha) = -2 + \frac{\sqrt{(a_n - c_n)^2 + 4b_n^2} + (a_n + c_n)}{\alpha}. \quad (\text{B.33})$$

$f(\alpha)$ is a decreasing function of α . In fact, if α is too small, then the root convergence factor $\rho = \frac{1}{2}f(\alpha) > 1$ and the algorithm will diverge (See Fig. B.2). Therefore, $\arg \min_{\alpha} f(\alpha) = \frac{1}{2}(a_n + c_n)$ when $\alpha \leq \frac{1}{2}(a_n + c_n)$.

Fig. B.2 shows an example of this 1D function $f(\alpha)$ for $a_n = 1, b_n = 1.1$ and $c_n = 4.3$. More generally, it has a unique minimum at $\alpha = \frac{1}{2}(a_n + c_n)$ for all $a_n c_n > b_n^2$. The solution to (B.30) is

$$\boxed{\alpha = \frac{1}{2}(a_n + c_n)}. \quad (\text{B.34})$$

The root convergence factor in this case is

$$\rho_1 = \frac{1}{2}f\left(\frac{1}{2}(a_n + c_n)\right) = \frac{\sqrt{(a_n - c_n)^2 + 4b_n^2}}{a_n + c_n}. \quad (\text{B.35})$$

One can verify that $\rho_1 > \rho_2$.

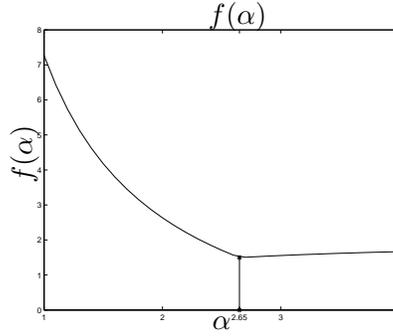


Figure B.2: $f(\alpha)$ for $a_n = 1$, $b_n = 1.1$ and $c_n = 4.3$. The minimum is at $\alpha = 2.65$.

B.4 Optimal Curvatures of Minimum Spectral Radius Criterion

We have derived sufficient conditions (B.15) and (B.16) for the majorization constraint (B.5) for separable quadratic surrogates. We have found the optimal diagonal curvature matrix (B.27) of the minimum spectral radius criterion without the majorization constraint. For the curvature matrix of constant times the identity matrix, a special case of diagonal curvature matrix, the corresponding results are given in (B.21) and (B.34). In this section, we put the constraints and optimal criterion together and derive optimal curvatures for separable quadratic surrogates.

B.4.1 Diagonal Curvature Matrix

We find the optimal curvature $\tilde{\mathbf{C}}_{\text{opt}}(\Psi, \mathbf{x}^{(n)}; \mathcal{C})$ satisfying the following definition:

1. Separable quadratic surrogates

$$\mathcal{C} = \left\{ \mathbf{C}_n = \mathbf{C}_n(\alpha_1, \alpha_2) = \begin{bmatrix} \alpha_1 & 0 \\ 0 & \alpha_2 \end{bmatrix} : \alpha_1, \alpha_2 \geq 0 \right\}.$$

2. \mathbf{C}_n satisfies the sufficient majorization condition, *i.e.*

$$\mathbf{C}_n \succeq \nabla^2 \Psi(\mathbf{x}), \quad \forall \mathbf{x}$$

$$\mathbf{C}_n \succeq \mathbf{0}.$$

3. The optimal criterion is

$$(\hat{\alpha}_1, \hat{\alpha}_2) = \arg \min_{\alpha_1, \alpha_2 \geq 0} \rho \left(\mathbf{I} - \mathbf{C}_n^{-1} \nabla^2 \Psi(\mathbf{x}^{(n)}) \right).$$

According to (B.15), (B.16) and (B.22), the minimization problem simplifies to

$$\begin{aligned} (\hat{\alpha}_1, \hat{\alpha}_2) &= \arg \min_{\alpha_1, \alpha_2 \geq 0} f(\alpha_1, \alpha_2), \\ f(\alpha_1, \alpha_2) &= \left| 2 - \frac{a_n}{\alpha_1} - \frac{c_n}{\alpha_2} \right| + \sqrt{\left(\frac{a_n}{\alpha_1} - \frac{c_n}{\alpha_2} \right)^2 + 4 \frac{b_n^2}{\alpha_1 \alpha_2}} \\ \text{s.t.} \quad &(\alpha_1 - a(\mathbf{x})) (\alpha_2 - c(\mathbf{x})) \geq b^2(\mathbf{x}), \quad \forall \mathbf{x} \\ &\alpha_1 + \alpha_2 \geq a(\mathbf{x}) + c(\mathbf{x}) \quad \forall \mathbf{x}. \end{aligned}$$

We have not found closed form solution to these conditions.

B.4.2 Curvature of Constant Times the Identity Matrix

We find the optimal curvature $\tilde{\mathbf{C}}_{\text{opt}}(\Psi, \mathbf{x}^{(n)}; \mathcal{C})$ satisfying the following definition:

1. Curvature matrix of constant times the identity matrix

$$\mathcal{C} = \{ \mathbf{C}_n = \mathbf{C}_n(\alpha) = \alpha \mathbf{I} : \alpha \geq 0 \}.$$

2. \mathbf{C}_n satisfies the sufficient majorization condition, *i.e.*

$$\mathbf{C}_n \succeq \nabla^2 \Psi(\mathbf{x}), \quad \forall \mathbf{x}$$

$$\mathbf{C}_n \succeq \mathbf{0}.$$

3. The optimal criterion is

$$\hat{\alpha} = \arg \min_{\alpha \geq 0} \rho \left(\mathbf{I} - \mathbf{C}_n^{-1} \nabla^2 \Psi(\mathbf{x}^{(n)}) \right).$$

According to (B.21) and (B.30), the minimization problem simplifies to

$$\begin{aligned}\hat{\alpha} &= \arg \min_{\alpha \geq 0} f(\alpha) \\ f(\alpha) &= \left| 2 - \frac{a_n + c_n}{\alpha} \right| + \frac{\sqrt{(a_n - c_n)^2 + 4b_n^2}}{\alpha} \\ \text{s.t.} \quad &\alpha \geq h(\mathbf{x}), \quad \forall \mathbf{x} \\ h(\mathbf{x}) &= \frac{1}{2}(a(\mathbf{x}) + c(\mathbf{x})) + \frac{1}{2}\sqrt{(a(\mathbf{x}) - c(\mathbf{x}))^2 + 4b^2(\mathbf{x})}.\end{aligned}$$

Since the majorization constraint (B.21) holds for all \mathbf{x} , we have $\alpha \geq \frac{1}{2}(a_n + c_n)$. According to the derivation in Section B.3.2, $f(\alpha)$ is an increasing function of α for this case. Therefore, the optimal curvature matrix of constant times the identity matrix is

$$\tilde{\mathbf{C}}_{\text{opt}} = \alpha_{\text{opt}} \mathbf{I} \quad (\text{B.36})$$

$$\alpha_{\text{opt}} = \max_{\mathbf{x}} \left\{ \frac{1}{2}(a(\mathbf{x}) + c(\mathbf{x})) + \frac{1}{2}\sqrt{(a(\mathbf{x}) - c(\mathbf{x}))^2 + 4b^2(\mathbf{x})} \right\}. \quad (\text{B.37})$$

The root convergence factor in this case is $\rho(\alpha_{\text{opt}}) = \frac{1}{2}f(\alpha_{\text{opt}})$.

In future work, it could be interesting to explore the difference between convergence using α_{opt} and using (B.34).

BIBLIOGRAPHY

BIBLIOGRAPHY

- [1] H. Erdoğan and J. A. Fessler. Monotonic algorithms for transmission tomography. *IEEE Trans. Med. Imag.*, 18(9):801–14, September 1999.
- [2] H. Erdoğan and J. A. Fessler. Ordered subsets algorithms for transmission tomography. *Phys. Med. Biol.*, 44(11):2835–51, November 1999.
- [3] M. Abella and J. A. Fessler. A new statistical image reconstruction algorithm for polyenergetic X-ray CT. In *Proc. IEEE Intl. Symp. Biomed. Imag.*, pages 165–8, 2009.
- [4] J. R. Adler, S. D. Chang, M. J. Murphy, J. Doty, P. Geis, and S. L. Hancock. The cyberknife: a frameless robotic system for radiosurgery. *Stereotact. Funct. Neurosurg.*, 69(1-4):124–8, 1997.
- [5] R. E. Alvarez and A. Macovski. Energy-selective reconstructions in X-ray computed tomography. *Phys. Med. Biol.*, 21(5):733–44, September 1976.
- [6] R. E. Alvarez and A. Macovski. X-ray spectral decomposition imaging system, 1977. U.S. Patent No. 4,029,963, June 14, 1977.
- [7] J. Ambrose. Computerized transverse axial scanning (tomography): Part II. Clinical application. *Brit. J. Radiology*, 46(552):1023–47, December 1973.
- [8] S. Basu and Y. Bresler. $O(N^3 \log(N))$ backprojection algorithm for the 3-D Radon transform. *IEEE Trans. Med. Imag.*, 21(2):76–88, February 2002.
- [9] S. Basu and B. De Man. Branchless distance driven projection and backprojection. In *Proc. SPIE 6065, Computational Imaging IV*, page 60650Y, 2006.
- [10] M. Bazalova, J.-F. Carrier, L. Beaulieu, and F. Verhaegen. Dual-energy CT-based material extraction for tissue segmentation in Monte Carlo dose calculations. *Phys. Med. Biol.*, 53(9):2439–56, May 2008.
- [11] R. I. Berbeco, S. B. Jiang, G. C. Sharp, G. T. Y. Chen, H. Mostafavi, and H. Shirato. Integrated radiotherapy imaging system (IRIS): design considerations of tumour tracking with linac gantry-mounted diagnostic X-ray systems with flat-panel detectors. *Phys. Med. Biol.*, 49(2):243–55, January 2004.

- [12] R. I. Berbeco, S. Nishioka, H. Shirato, G. T. Y. Chen, and S. B. Jiang. Residual motion of lung tumours in gated radiotherapy with external respiratory surrogates. *Phys. Med. Biol.*, 50(16):3655–68, August 2005.
- [13] R. Billingsley. *Probability and measure*. Wiley, New York, 2 edition, 1986.
- [14] W. Birkfellner, R. Seemann, M. Figl, J. Hummel, C. Ede, P. Homolka, X. Yang, P. Niederer, and H. Bergmann. Wobbled splatting—a fast perspective volume rendering method for simulation of x-ray images from CT. *Phys. Med. Biol.*, 50(9):N73–84, May 2005.
- [15] P. Boccacci, P. Bonetto, P. Calvini, and A. R. Formiconi. A simple model for the efficient correction of collimator blur in 3D SPECT imaging. *Inverse Prob.*, 15(4):907–30, August 1999.
- [16] J. Brokish and Y. Bresler. A hierarchical algorithm for fast backprojection in helical cone-beam tomography. In *Proc. IEEE Intl. Symp. Biomed. Imag.*, page 1420, 2004.
- [17] J. Brokish and Y. Bresler. Ultra-fast hierarchical backprojection for micro-ct reconstruction. In *Proc. IEEE Intl. Symp. Biomed. Imag.*, pages 4460–3, 2007.
- [18] R. A. Brooks and G. Di Chiro. Beam-hardening in X-ray reconstructive tomography. *Phys. Med. Biol.*, 21(3):390–8, May 1976.
- [19] J. A. Browne, J. M. Boone, and T. J. Holmes. Maximum-likelihood X-ray computed-tomography finite-beamwidth considerations. *Appl. Optics*, 34(23):5199–209, August 1995.
- [20] H. N. Cardinal and A. Fenster. An accurate method for direct dual-energy calibration and decomposition. *Med. Phys.*, 17(3):327–41, May 1990.
- [21] P. S. Cho, R. H. Johnson, and T. W. Griffin. Cone-beam CT for radiotherapy applications. *Phys. Med. Biol.*, 40(11):1863–1883, November 1995.
- [22] Z. H. Cho, C. M. Chen, and S. Y. Lee. Incremental algorithm—A new fast backprojection scheme for parallel geometries. *IEEE Trans. Med. Imag.*, 9(2):207–17, June 1990.
- [23] S. Y. Chun and J. A. Fessler. A simple regularizer for B-spline nonrigid image registration that encourages local invertibility. *IEEE J. Sel. Top. Sig. Proc.*, 3(1):159–69, February 2009. Special Issue on Digital Image Processing Techniques for Oncology.
- [24] N. H. Clinthorne. A constrained dual-energy reconstruction method for material-selective transmission tomography. *Nucl. Instr. Meth. Phys. Res. A.*, 353(1):347–8, December 1994.
- [25] A. P. Colijn and F. J. Beekman. Accelerated simulation of cone beam X-ray scatter projections. *IEEE Trans. Med. Imag.*, 23(5):584–90, May 2004.

- [26] B. De Man and S. Basu. Distance-driven projection and backprojection in three dimensions. *Phys. Med. Biol.*, 49(11):2463–75, June 2004.
- [27] A. R. De Pierro. A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography. *IEEE Trans. Med. Imag.*, 14(1):132–7, March 1995.
- [28] I. A. Elbakri and J. A. Fessler. Statistical image reconstruction for polyenergetic X-ray computed tomography. *IEEE Trans. Med. Imag.*, 21(2):89–99, February 2002.
- [29] I. A. Elbakri and J. A. Fessler. Segmentation-free statistical image reconstruction for polyenergetic X-ray computed tomography with experimental validation. *Phys. Med. Biol.*, 48(15):2543–78, August 2003.
- [30] L. A. Feldkamp, L. C. Davis, and J. W. Kress. Practical cone beam algorithm. *J. Opt. Soc. Am. A*, 1(6):612–9, June 1984.
- [31] J. A. Fessler. Penalized weighted least-squares image reconstruction for positron emission tomography. *IEEE Trans. Med. Imag.*, 13(2):290–300, June 1994.
- [32] 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.
- [33] 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.
- [34] J. A. Fessler. Statistical image reconstruction methods for transmission tomography. In M. Sonka and J. Michael Fitzpatrick, editors, *Handbook of Medical Imaging, Volume 2. Medical Image Processing and Analysis*, pages 1–70. SPIE, Bellingham, 2000.
- [35] J. A. Fessler. Method for statistically reconstructing images from a plurality of transmission measurements having energy diversity and image reconstructor apparatus utilizing the method, 2004. US Patent 6,754,298.
- [36] 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.
- [37] J. A. Fessler, I. Elbakri, P. Sukovic, and N. H. Clinthorne. Maximum-likelihood dual-energy tomographic image reconstruction. In *Proc. SPIE 4684, Medical Imaging 2002: Image Proc.*, volume 1, pages 38–49, 2002.
- [38] R. R. Galigekere, K. Wiesent, and D. W. Holdsworth. Cone-beam reprojection using projection-matrices. *IEEE Trans. Med. Imag.*, 22(10):1202–14, October 2003.

- [39] G. C. Giakos, S. Chowdhury, N. Shah, K. Mehta, S. Sumrain, A. Passerini, N. Patnekar, E. A. Evans, L. Fraiwan, O. C. Ugweje, and R. Nemer. Signal evaluation of a novel dual-energy multimedia imaging sensor. *IEEE Trans. Med. Imag.*, 51(5):949–54, October 2002.
- [40] G. H. Glover and N. J. Pelc. Nonlinear partial volume artifacts in x-ray computed tomography. *Med. Phys.*, 7(3):238–48, May 1980.
- [41] D. J. Godfrey, F-F. Yin, M. Oldham, S. Yoo, and C. Willett. Digital tomosynthesis with an on-board kilovoltage imaging device. *Int. J. Radiat. Oncol. Biol. Phys.*, 65(1):8–15, May 2006.
- [42] G. Harikumar and Y. Bresler. Exact image deconvolution from multiple FIR blurs. *IEEE Trans. Im. Proc.*, 8(6):846–62, June 1999.
- [43] G. T. Herman. Correction for beam hardening in computed tomography. *Phys. Med. Biol.*, 24(1):81–106, January 1979.
- [44] D. L. G. Hill, P. G. Batchelor, M. Holden, and D. J. Hawkes. Medical image registration. *Phys. Med. Biol.*, 46(3):R1–47, March 2001.
- [45] I. K. Hong, S. T. Chung, H. K. Kim, Y. B. Kim, Y. D. Son, and Z. H. Cho. Ultra fast symmetry and SIMD-based projection-backprojection (SSP) algorithm for 3-D PET image reconstruction. *IEEE Trans. Med. Imag.*, 26(6):789–803, June 2007.
- [46] R. A. Horn and C. R. Johnson. *Matrix analysis*. Cambridge Univ. Press, Cambridge, 1985.
- [47] H. M. Hudson and R. S. Larkin. Accelerated image reconstruction using ordered subsets of projection data. *IEEE Trans. Med. Imag.*, 13(4):601–9, December 1994.
- [48] W. Huh and J. A. Fessler. Model-based image reconstruction for dual-energy X-ray CT with fast kVp switching. In *Proc. IEEE Intl. Symp. Biomed. Imag.*, pages 326–9, 2009.
- [49] W. Huh and J. A. Fessler. Iterative image reconstruction for dual-energy x-ray CT using regularized material sinogram estimates. In *Proc. IEEE Intl. Symp. Biomed. Imag.*, 2011. To appear as oral 1167.
- [50] W. Huh, J. A. Fessler, A. M. Alessio, and P. E. Kinahan. Fast kVp-switching dual energy CT for PET attenuation correction. In *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, pages 2510–5, 2009.
- [51] D. R. Hunter and K. Lange. A tutorial on MM algorithms. *American Statistician*, 58(1):30–7, February 2004.
- [52] Imaginis. History of radiation therapy. World Wide Web address http://www.imaginis.com/radiotherapy/radio_history.asp.

- [53] M. Jacobson. *Approaches to motion-corrected PET image reconstruction from respiratory gated projection data*. PhD thesis, Univ. of Michigan, Ann Arbor, MI, 48109-2122, Ann Arbor, MI, 2006.
- [54] D. A. Jaffray, D. G. Drakes, M. Moreau, A. A. Martinez, and J. W. Wong. A radiographic and tomographic imaging system integrated into a medical linear accelerator for localization of bone and soft-tissue targets. *Int. J. Radiat. Oncol. Biol. Phys.*, 45(3):773–89, October 1999.
- [55] D. A. Jaffray, J. H. Siewerdsen, J. W. Wong, and A. A. Martinez. Flat-panel cone-beam computed tomography for image-guided radiation therapy. *Int. J. Radiat. Oncol. Biol. Phys.*, 53(5):1337–49, August 2002.
- [56] P. M. Joseph. An improved algorithm for reprojecting rays through pixel images. *IEEE Trans. Med. Imag.*, 1(3):192–6, November 1982.
- [57] P. M. Joseph and R. D. Spital. A method for correcting bone induced artifacts in computed tomography scanners. *J. Comp. Assisted Tomo.*, 2(1):100–8, January 1978.
- [58] P. M. Joseph and R. D. Spital. The exponential edge-gradient effect in x-ray computed tomography. *Phys. Med. Biol.*, 26(3):473–87, May 1981.
- [59] M. Kachelriess, M. Knaup, and O. Bockenbach. Hyperfast parallel-beam and cone-beam backprojection using the cell general purpose hardware. *Med. Phys.*, 34(4):1474–86, April 2007.
- [60] A. C. Kak and M. Slaney. *Principles of computerized tomographic imaging*. IEEE Press, New York, 1988.
- [61] S. Kappler, M. Grasruck, D. Niederlöhner, M. Strassburg, and S. Wirth. Dual-energy performance of dual-kVp in comparison to dual-layer and quantum-counting CT system concepts. In *Proc. SPIE 7258, Medical Imaging 2009: Phys. Med. Im.*, page 725842, 2009.
- [62] J. Kim. *Intensity based image registration using robust similarity measure and constrained optimization: applications for radiation therapy*. PhD thesis, Univ. of Michigan, Ann Arbor, MI, 48109-2122, Ann Arbor, MI, 2004.
- [63] J. Kim, J. A. Fessler, K. L. Lam, J. M. Balter, and R. K. Ten Haken. A feasibility study of mutual information based set-up error estimator for radiotherapy. *Med. Phys.*, 28(12):2507–17, December 2001.
- [64] J. K. Kim, J. A. Fessler, and Z. Zhang. FPGA implementation of forward-projection for X-ray CT using separable footprints. In *UM Graduate Symposium*, 2010.
- [65] J-K. Kim, Z. Zhang, and J. A. Fessler. Hardware acceleration of iterative image reconstruction for X-ray computed tomography. In *Proc. IEEE Conf. Acoust. Speech Sig. Proc.*, 2011. To appear.

- [66] P. E. Kinahan, A. M. Alessio, and J. A. Fessler. Dual energy CT attenuation correction methods for quantitative assessment of response to cancer therapy with PET/CT imaging. *Technology in Cancer Research and Treatment*, 5(4):319–28, August 2006.
- [67] Z. Kolitsi, G. Panayiotakis, V. Anastassopoulos, A. Scodras, and N. Pallikarakis. A multiple projection method for digital tomosynthesis. *Med. Phys.*, 19(4):1045–50, July 1992.
- [68] P. O. Kotzki, D. Mariano-Goulart, and M. Rossi. Prototype of dual energy X-ray tomodesimeter for lumbar spine bone mineral density measurements; choice of the reconstruction algorithm and first experimental results. *Phys. Med. Biol.*, 37(12):2253–65, December 1992.
- [69] Y. Kyriakou, E. Meyer, D. Prell, and M. Kachelriess. Empirical beam hardening correction (EBHC) for CT. *Med. Phys.*, 37(10):5179–, October 2010.
- [70] D. S. Lalush and B. M. W. Tsui. A generalized Gibbs prior for maximum a posteriori reconstruction in SPECT. *Phys. Med. Biol.*, 38(6):729–41, June 1993.
- [71] K. Lange and J. A. Fessler. Globally convergent algorithms for maximum a posteriori transmission tomography. *IEEE Trans. Im. Proc.*, 4(10):1430–8, October 1995.
- [72] K. Lange, D. R. Hunter, and I. Yang. Optimization transfer using surrogate objective functions. *J. Computational and Graphical Stat.*, 9(1):1–20, March 2000.
- [73] P. J. LaRiviere and X. Pan. Sampling and aliasing consequences of quarter-detector offset use in helical CT. *IEEE Trans. Med. Imag.*, 23(6):738–49, June 2004.
- [74] R. M. Leahy and J. Qi. Statistical approaches in quantitative positron emission tomography. *Statistics and Computing*, 10(2):147–65, April 2000.
- [75] L. A. Lehmann and R. E. Alvarez. Energy-selective radiography: A review. In C Orton J Kareiakas, S Thomas, editor, *Digital radiography: Selected topics*, pages 145–88. Plenum, New York, 1986.
- [76] L. Lemieux, R. Jagoe, D. R. Fish, N. D. Kitchen, and D. G. T. Thomas. A patient-to-computed-tomography image registration method based on digitally reconstructed radiographs. *Med. Phys.*, 21(11):1749–60, November 1994.
- [77] T. Li, E. Schreibmann, Y. Yang, and L. Xing. Motion correction for improved target localization with on-board cone-beam computed tomography. *Phys. Med. Biol.*, 51(2):253–68, January 2006.
- [78] X. Liu, L. Yu, A. N. Primak, and C. H. McCollough. Quantitative imaging of element composition and mass fraction using dual-energy CT: Three-material decomposition. *Med. Phys.*, 36(5):1602–9, April 2009.

- [79] Y. Long, J. A. Fessler, and J. M. Balter. A 3D forward and back-projection method for X-ray CT using separable footprint. In *Proc. Intl. Mtg. on Fully 3D Image Recon. in Rad. and Nuc. Med.*, pages 146–9, 2009. Winner of poster award.
- [80] Y. Long, J. A. Fessler, and J. M. Balter. 3D forward and back-projection for X-ray CT using separable footprints. *IEEE Trans. Med. Imag.*, 29(11):1839–50, November 2010.
- [81] Y. Long, J. A. Fessler, and J. M. Balter. 3D forward and back-projection for X-ray CT using separable footprints with trapezoid functions. In *Proc. First Intl. Mtg. on image formation in X-ray computed tomography*, pages 216–9, 2010.
- [82] Y. Long, J. A. Fessler, and J. M. Balter. Accuracy estimation for projection-to-volume targeting during rotational therapy: A feasibility study. *Med. Phys.*, 37(6):2480–90, June 2010.
- [83] Y. Long, J. A. Fessler, and J. M. Balter. Accuracy limits for projection-to-volume targeting during arc therapy. In *Proc. Amer. Assoc. Phys. Med.*, page 3154, 2010.
- [84] Y. Long, J. A. Fessler, and J. M. Balter. Fast and accurate three-dimensional forward and back-projection methods, 2010. Provisional patent application.
- [85] Y. Long, J. A. Fessler, and J. M. Balter. Two-material decomposition from a single CT scan using statistical image reconstruction. In *Proc. Amer. Assoc. Phys. Med.*, 2011. To appear 15164. Wed. 11:20 AM, WC-C-110-6 oral.
- [86] C. Maass, R. Grimmer, and M. Kachelriess. Dual energy CT material decomposition from inconsistent rays (MDIR). In *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, pages 3446–3452, 2009.
- [87] A. Macovski. *Medical imaging systems*. Prentice-Hall, New Jersey, 1983.
- [88] A. Macovski, R. E. Alvarez, J. Chan, J. P. Stonestrom, and L. M. Zatz. Energy dependent reconstruction in X-ray computerized tomography. *Computers in Biology and Medicine*, 6(4):325–36, October 1976.
- [89] M. Maddah, P. Mendonca, and R. Bhotika. Physically meaningful virtual unenhanced image reconstruction from dual-energy ct. In *Proc. IEEE Intl. Symp. Biomed. Imag.*, pages 808–, 2010.
- [90] J. S. Maltz, F. Sprenger, J. Fuerst, A. Paidi, F. Fadler, and A. R. Bani-Hashemi. Fixed gantry tomosynthesis system for radiation therapy image guidance based on a multiple source x-ray tube with carbon nanotube cathodes. *Med. Phys.*, 36(5):1624–36, May 2009.
- [91] C. Markham and J. Fryar. Element specific imaging in computerised tomography using a tube source of X-rays and a low energy-resolution detector system. *Nucl. Instr. Meth. Phys. Res. A.*, 324(1-2):383–8, January 1993.

- [92] W. H. Marshall, R. E. Alvarez, and A. Macovski. Initial results with prereconstruction dual-energy computed tomography (PREDECT). *Radiology*, 140(2):421–30, August 1981.
- [93] S. Matej and R. M. Lewitt. Practical considerations for 3-D image reconstruction using spherically symmetric volume elements. *IEEE Trans. Med. Imag.*, 15(1):68–78, February 1996.
- [94] S. Matej and R. M. Lewitt. 3D-FRP: direct Fourier reconstruction with Fourier reprojection for fully 3-D PET. *IEEE Trans. Nuc. Sci.*, 48(4-2):78–1385, August 2001.
- [95] J. Maurer, D. Godfrey, Z. Wang, and F-F. Yin. On-board four-dimensional digital tomosynthesis: First experimental results. *Med. Phys.*, 35(8):3574–3583, August 2008.
- [96] P. R. S. Mendonca, R. Bhotika, B. W. Thomsen, P. E. Licato, and M. C. Joshi. Multi-material decomposition of dual-energy CT data. In *spie*, volume 7622, page 76221W, 2010.
- [97] G. J. Michael. Tissue analysis using dual energy CT. *Australasian Physical & Engineering Sciences in Medicine*, 15(1):25–37, March 1992.
- [98] M. A. Mosleh-Shirazi, P. M. Evans, W. Swindell, S. Webb, and M. Partridge. A cone-beam megavoltage CT scanner for treatment verification in conformal radiotherapy. *Radiother. Oncol.*, 48(3):319–28, September 1998.
- [99] K. Mueller and R. Yagel. Rapid 3-D cone-beam reconstruction with the simultaneous algebraic reconstruction technique (SART) using 2-D texture mapping hardware. *IEEE Trans. Med. Imag.*, 19(12):1227–37, December 2000.
- [100] K. Mueller, R. Yagel, and J. J. Wheeler. Anti-aliased three-dimensional cone-beam reconstruction of low-contrast objects with algebraic methods. *IEEE Trans. Med. Imag.*, 18(6):519–37, June 1999.
- [101] K. Mueller, R. Yagel, and J. J. Wheller. Fast implementations of algebraic methods for three-dimensional reconstruction from cone-beam data. *IEEE Trans. Med. Imag.*, 18(6):538–48, June 1999.
- [102] F. Natterer. Sampling in fan beam tomography. *SIAM J. Appl. Math.*, 53(2):358–80, April 1993.
- [103] L. P. Nogueira, R. C. Barroso, C. J. G. Pinheiro, D. Braz, L. F. de Oliveira, G. Tromba, and N. Sodini. Mapping lead distribution in bones by dual-energy computed microtomography with synchrotron radiation. In *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, pages 3471–4, 2009.
- [104] J. Noh, J. A. Fessler, and P. E. Kinahan. Statistical sinogram restoration in dual-energy CT for PET attenuation correction. *IEEE Trans. Med. Imag.*, 28(11):1688–702, November 2009.

- [105] K. Otto. Volumetric modulated arc therapy: IMRT in a single gantry arc. *Med. Phys.*, 35(1):310–17, January 2008.
- [106] G. Pang, A. Bani-Hashemi, P. Au, P. F. O’Brien, J. A. Rowlands, G. Morton, T. Lim, P. Cheung, and A. Loblaw. Megavoltage cone beam digital tomosynthesis (MV-CBDT) for image-guided radiotherapy: a clinical investigational system. *Phys. Med. Biol.*, 53(4):999–1013, January 2008.
- [107] G. P. Penney, J. Weese, J. A. Little, P. Desmedt, D. L. G. Hill, and D. J. Hawkes. A comparison of similarity measures for use in 2-D–3-D medical image registration. *IEEE Trans. Med. Imag.*, 17(4):586–95, August 1998.
- [108] T. M. Peters. Algorithm for fast back- and reprojection in computed tomography. *IEEE Trans. Nuc. Sci.*, 28(4):3641–7, August 1981.
- [109] J. Pouliot, A. Bani-Hashemi, J. Chen, M. Svatos, F. Ghelmansarai, M. Mitschke, M. Aubin, P. Xia, O. Morin, K. Bucci, and M. Roach. Low-dose megavoltage cone-beam CT for radiation therapy. *Int. J. Radiat. Oncol. Biol. Phys.*, 61(2):552–60, February 2005.
- [110] G. Prax, G. Chinn, P. D. Olcott, and C. S. Levin. Fast, accurate and shift-varying line projections for iterative reconstruction using the GPU. *IEEE Trans. Med. Imag.*, 28(3):435–45, March 2009.
- [111] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. *Numerical recipes in C*. Cambridge Univ. Press, New York, 2 edition, 1992.
- [112] A. N. Primak, J. C. Ramirez Giraldo, X. Liu, L. Yu, and C. H. McCollough. Improved dual-energy material discrimination for dual-source CT by means of additional spectral filtration. *Med. Phys.*, 36(4):1359–69, March 2009.
- [113] RadiologyInfo. Intensity-Modulated radiation therapy (IMRT). World Wide Web address <http://www.radiologyinfo.org/en/info.cfm?PG=imrt>.
- [114] L. Ren, D. J. Godfrey, H. Yan, Q. J. Wu, and F-F. Yin. Automatic registration between reference and on-board digital tomosynthesis images for positioning verification. *Med. Phys.*, 35(2):664–672, February 2008.
- [115] L. Ren, J. Zhang, D. Thongphiew, D. J. Godfrey, Q. J. Wu, S-M. Zhou, and F-F. Yin. A novel digital tomosynthesis (DTS) reconstruction method using a deformation field map. *Med. Phys.*, 35(7):3110–5, July 2008.
- [116] Cancer research UK. Scans for radiotherapy. World Wide Web address <http://www.cancerhelp.org.uk/help/default.asp?page=3604>.
- [117] Cancer research UK. What is image guided radiotherapy? World Wide Web address <http://www.cancerhelp.org.uk/about-cancer/cancer-questions/what-is-image-guided-radiotherapy>.

- [118] Cancer research UK. What radiotherapy is. World Wide Web address <http://www.cancerhelp.org.uk/help/default.asp?page=3586>.
- [119] Cancer research UK. Your radiotherapy treatment plan. World Wide Web address <http://www.cancerhelp.org.uk/help/default.asp?page=3588>.
- [120] C. Riddell and Y. Trouset. Rectification for cone-beam projection and backprojection. *IEEE Trans. Med. Imag.*, 25(7):950–62, July 2006.
- [121] R. T. Ritchings and B. R. Pullan. A technique for simultaneous dual energy scanning. *J. Comp. Assisted Tomo.*, 3(6):842–6, December 1979.
- [122] C. Ruth and P. M. Joseph. Estimation of a photon energy spectrum for a computed tomography scanner. *Med. Phys.*, 24(5):695–702, May 1997.
- [123] B. Rutt and A. Fenster. Split-filter computed tomography: A simple technique for dual energy scanning. *J. Comp. Assisted Tomo.*, 4(4):501–9, August 1980.
- [124] V. Sarkar, C. Shi, P. Rassiah-Szegedi, T. Eng, A. Diaz, and N. Papanikolaou. A feasibility study on the use of digital tomosynthesis with individually-acquired megavoltage portal images for target localization. *J. BUON*, 14(1):103–8, January 2009.
- [125] K. Sauer and C. Bouman. A local update strategy for iterative reconstruction from projections. *IEEE Trans. Sig. Proc.*, 41(2):534–48, February 1993.
- [126] J. J. Scheins, F. Boschen, and H. Herzog. Analytical calculation of volumes-of-intersection for iterative, fully 3-D PET reconstruction. *IEEE Trans. Med. Imag.*, 25(10):1363–9, October 2006.
- [127] W. Schneider, T. Bortfeld, and W. Schlegel. Correlation between CT numbers and tissue parameters needed for Monte Carlo simulations of clinical dose distribution. *Phys. Med. Biol.*, 45(2):459–78, February 2000.
- [128] C. Schretter. A fast tube of response ray-tracer. *Med. Phys.*, 33(12):4744–8, December 2006.
- [129] W. P. Segars and B. M. W. Tsui. Study of the efficacy of respiratory gating in myocardial SPECT using the new 4-D NCAT phantom. *IEEE Trans. Nuc. Sci.*, 49(3):675–9, June 2002.
- [130] R. L. Siddon. Fast calculation of the exact radiological path for a three-dimensional CT array. *Med. Phys.*, 12(2):252–5, March 1985.
- [131] D. L. Snyder, C. W. Helstrom, A. D. Lanterman, M. Faisal, and R. L. White. Compensation for readout noise in CCD images. *J. Opt. Soc. Am. A*, 12(2):272–83, February 1995.
- [132] D. L. Snyder and M. I. Miller. The use of sieves to stabilize images produced with the EM algorithm for emission tomography. *IEEE Trans. Nuc. Sci.*, 32(5):3864–71, October 1985.

- [133] K. Sourbelle, M. Kachelriess, and W. A. Kalender. Empirical water precorrection for cone-beam computed tomography. In *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, volume 4, pages 1871–5, 2005.
- [134] J. Spoerk, H. Bergmann, F. Wanschitz, S. Dong, and W. Birkfellner. Fast DRR splat rendering using common consumer graphics hardware. *Med. Phys.*, 34(11):4302–8, November 2007.
- [135] S. Srivastava and J. A. Fessler. Simplified statistical image reconstruction algorithm for polyenergetic X-ray CT. In *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, volume 3, pages 1551–5, 2005.
- [136] 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.
- [137] 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.
- [138] C. W. Stevens, R. F. Munden, K. M. Forster, J. F. Kelly, Z. Liao, G. Starkschall, S. Tucker, and R. Komaki. Respiratory-driven lung tumor motion is independent of tumor size, tumor location, and pulmonary function. *Int. J. Radiat. Oncol. Biol. Phys.*, 51(1):62–8, September 2001.
- [139] J. P. Stonestrom, R. E. Alvarez, and A. Macovski. A framework for spectral artifact corrections in X-ray CT. *IEEE Trans. Biomed. Engin.*, 28(2):128–41, February 1981.
- [140] J. Stutzel, U. Oelfke, and S. Nill. A quantitative image quality comparison of four different image guided radiotherapy devices. *Radiother. Oncol.*, 86(1):20–4, January 2008.
- [141] P. Sukovic and N. H. Clinthorne. Data weighted vs. non-data weighted dual energy reconstructions for X-ray tomography. In *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, volume 3, pages 1481–3, 1998.
- [142] P. Sukovic and N. H. Clinthorne. Basis material decomposition using triple-energy X-ray computed tomography. In *IEEE Instrumentation and Measurement Technology Conference, Venice*, volume 3, pages 1615–8, 1999.
- [143] P. Sukovic and N. H. Clinthorne. Design of an experimental system for dual energy X-ray CT. In *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, volume 2, pages 1021–2, 1999.
- [144] P. Sukovic and N. H. Clinthorne. Penalized weighted least-squares image reconstruction for dual energy X-ray transmission tomography. *IEEE Trans. Med. Imag.*, 19(11):1075–81, November 2000.

- [145] J. Sunnegårdh and P-E. Danielsson. A new anti-aliased projection operator for iterative CT reconstruction. In *Proc. Intl. Mtg. on Fully 3D Image Recon. in Rad. and Nuc. Med*, pages 124–7, 2007.
- [146] R. Taschereau, R. W. Silverman, and A. F. Chatziioannou. Dual-energy attenuation coefficient decomposition with differential filtration and application to a microCT scanner. *Phys. Med. Biol.*, 55(4):1141–1155, February 2010.
- [147] 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.
- [148] K. Todros and J. Tabrikian. General classes of performance lower bounds for parameter estimation—Part I: Non-Bayesian bounds for unbiased estimators. *IEEE Trans. Info. Theory*, 56(10):5045–63, October 2010.
- [149] K. Todros and J. Tabrikian. General classes of performance lower bounds for parameter estimation: Part II: Bayesian bounds. *IEEE Trans. Info. Theory*, 56(10):5064–82, October 2010.
- [150] I. B. Tutar, R. Managuli, V. Shamdasani, P. S. Cho, S. D. Pathak, and Y. Kim. Tomosynthesis-based localization of radioactive seeds in prostate brachytherapy. *Med. Phys.*, 30(12):3135–42, December 2003.
- [151] M. Unser. Splines: A perfect fit for signal and image processing. *spmag*, 16(6):22–38, November 1999.
- [152] M. Unser, A. Aldroubi, and M. Eden. The L_2 polynomial spline pyramid. *IEEE Trans. Patt. Anal. Mach. Int.*, 15(4):364–79, April 1993.
- [153] L. Westover. Footprint evaluation for volume rendering. In *Intl. Conf. on Computer Graphics and Interactive Techniques*, pages 367–76, 1990.
- [154] M. Wu and J. A. Fessler. GPU acceleration of 3D forward and backward projection using separable footprints for X-ray CT image reconstruction. In *Proc. Intl. Mtg. on Fully 3D Image Recon. in Rad. and Nuc. Med*, 2011.
- [155] Q. J. Wu, D. J. Godfrey, Z. Wang, J. Zhang, S. Zhou, S. Yoo, D. M. Brizel, and F-F. Yin. On-board patient positioning for head-and-neck IMRT: Comparing digital tomosynthesis to kilovoltage radiography and cone-beam computed tomography. *Int. J. Radiat. Oncol. Biol. Phys.*, 69(2):596–606, 2007.
- [156] M. Yavuz and J. A. Fessler. Statistical image reconstruction methods for randoms-precorrected PET scans. *Med. Im. Anal.*, 2(4):369–78, December 1998.
- [157] M. Yavuz and J. A. Fessler. Penalized-likelihood estimators and noise analysis for randoms-precorrected PET transmission scans. *IEEE Trans. Med. Imag.*, 18(8):665–74, August 1999.

- [158] S. Yoo, Q. J. Wu, D. Godfrey, H. Yan, L. Ren, S. Das, W. R. Lee, and F-F. Yin. Clinical evaluation of positioning verification using digital tomosynthesis and bony anatomy and soft tissues for prostate image-guided radiotherapy. *Int. J. Radiat. Oncol. Biol. Phys.*, 73(1):296–305, 2009.
- [159] C. X. Yu. Intensity-modulated arc therapy with dynamic multileaf collimation: An alternative to tomotherapy. *Phys. Med. Biol.*, 40(9):1435–49, September 1995.
- [160] L. Yu, X. Liu, and C. H. McCollough. Pre-reconstruction three-material decomposition in dual-energy CT. In *Proc. SPIE 7258, Medical Imaging 2009: Phys. Med. Im.*, page 72581V, 2009.
- [161] W. Zbijewski and F. J. Beekman. Characterization and suppression of edge and aliasing artefacts in iterative x-ray CT reconstruction. *Phys. Med. Biol.*, 49(1):145–158, January 2004.
- [162] W. Zbijewski and F. J. Beekman. Comparison of methods for suppressing edge and aliasing artefacts in iterative x-ray CT reconstruction. *Phys. Med. Biol.*, 51(7):1877–90, April 2006.
- [163] R. Zeng, J. A. Fessler, and J. M. Balter. Respiratory motion estimation from slowly rotating X-ray projections: Theory and simulation. *Med. Phys.*, 32(4):984–91, April 2005.
- [164] R. Zeng, J. A. Fessler, and J. M. Balter. Estimating 3-D respiratory motion from orbiting views by tomographic image registration. *IEEE Trans. Med. Imag.*, 26(2):153–63, February 2007.
- [165] Y. Zhang-O’Connor and J. A. Fessler. Fourier-based forward and back-projectors in iterative fan-beam tomographic image reconstruction. *IEEE Trans. Med. Imag.*, 25(5):582–9, May 2006.
- [166] H. Zhao and A. J. Reader. Fast ray-tracing technique to calculate line integral paths in voxel arrays. In *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, pages 2808–12, 2003.
- [167] A. Ziegler, T. Köhler, T. Nielsen, and R. Proksa. Efficient projection and backprojection scheme for spherically symmetric basis functions in divergent beam geometry. *Med. Phys.*, 33(12):4653–63, December 2006.
- [168] A. Ziegler, T. Nielsen, and M. Grass. Iterative reconstruction of a region of interest for transmission tomography. *Med. Phys.*, 35(4):1317–27, April 2008.
- [169] B. Zitová and J. Flusser. Image registration methods: a survey. *Im. and Vision Computing*, 21(11):977–1000, October 2003.