Summer 8-15-2019

Multi-dimensional extension of the alternating minimization algorithm in x-ray computed tomography

Jingwei Lu
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS

McKelvey School of Engineering

Department of Electrical & Systems Engineering

Dissertation Examination Committee:
Joseph A. O'Sullivan, Chair
Ulugbek Kamilov
Matthew Lew
David G. Politte
Umberto Villa

Multi-Dimensional Extension of the Alternating Minimization Algorithm in X-ray Computed Tomography

by

Jingwei Lu

A dissertation presented to
The Graduate School
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

August 2019
Saint Louis, Missouri

Contents

List of Figures

List of Tables

Acknowledgments

First of all, my sincerest thanks to my parents, Linlin Lu and Ping Guan, and my dear Zhengzhen Wang, for the days you have been with me.

I would like to thank my advisor, Dr. Joseph A. O'Sullivan. His guidance, experience, and knowledge inspired me through my doctoral research and made me a complete doctor. I would like to express my great thanks to his trust and appreciation of my work.

I would like to thank my committee members, Dr. David G. Politte, Dr. Matthew Lew, Dr. Ulugbek Kamilov, and Dr. Umberto Villa for their time and suggestions that helped me complete this dissertation.

I would like to thank my collaborators, Dr. Shuangyue Zhang, Dr. David G. Politte, Dr. Bruce Whiting, Dr. Jeffery F. Williamson, Dr. Dong Han, Dr. Tianyu Zhao, and Carl Bosch for their help during my research projects.

I would like to thank my labmates, including but not limited to Dr. Linyun He, Dr. Soysal Degirmenci, Dr. Ayan Mitra, Dr. Jianyong Jiang, Dr. Homayoon Ranjbar, Dr. Yaqi Chen,

Dr. Ke Li, Dr. Ikenna C. Odinaka, Tao Ge, Rui Liao, Tyler B. Webb, and Maria J. Medrano for their help with my research.

I would like to thank the funding support from NIH 1R01CA149305, NIH 1R01CA212638, and NIH 5T32EB01485505.

Jingwei Lu

Washington University in Saint Louis

August 2019

Dedicated to my parents and Zhengzhen

ABSTRACT OF THE DISSERTATION

Multi-Dimensional Extension of the Alternating Minimization Algorithm in X-ray Computed Tomography

by

Jingwei Lu

Doctor of Philosophy in Electrical Engineering

Washington University in St. Louis, 2019

Professor Joseph A. O'Sullivan, Chair

X-ray computed tomography (CT) is an important and effective tool in medical and industrial imaging applications. The state-of-the-art methods to reconstruct CT images have had great development but also face challenges. This dissertation derives novel algorithms to reduce bias and metal artifacts in a wide variety of imaging modalities and increase performance in low-dose scenarios.

The most widely available CT systems still use the single-energy CT (SECT), which is good at showing the anatomic structure of the patient body. However, in SECT image reconstruction, energy-related information is lost. In applications like radiation treatment planning and dose prediction, accurate energy-related information is needed. Spectral CT has shown the potential to extract energy-related information.

Dual-energy CT (DECT) is the first successful implementation of spectral CT. By using two different spectra, the energy-related information can be exported by reconstructing basis-material images. A sinogram-based decomposition method has shown good performance in clinical applications. However, when the x-ray dose level is low, the sinogram-based decomposition methods generate biased estimates. The bias increases rapidly when the dose level decreases. The bias comes from the ill-posed statistical model in the sinogram-decomposition method. To eliminate the bias in low-dose cases, a joint statistical image reconstruction (JSIR) method using the dual-energy alternating minimization (DEAM) algorithm is proposed. By correcting the ill-posed statistical model, a relative error as high as 15% in the sinogram-based decomposition method can be reduced to less than 1% with DEAM, which is an approximately unbiased estimation.

Photon counting CT (PCCT) is an emerging CT technique that also can resolve the energy information. By using photon counting detectors (PCD), PCCT keeps track of the energy of every photon received. Though PCDs have an entirely different physical performance from the energy-integrating detectors used in DECT, the problem of biased estimation with the sinogram-decomposition method remains. Based on DEAM, a multi-energy alternating minimization (MEAM) algorithm for PCCT is proposed. In the simulation experiments, MEAM can effectively reduce the bias by more than 90%.

Metal artifacts have been a concern since x-ray CT came into medical imaging. When there exist dense or metal materials in the scanned object, the image quality may suffer severe artifacts. The auxiliary sinogram alternating minimization (ASAM) algorithm is proposed to take advantages of two major categories of methods to deal with metal artifacts: the pre-processing method and statistical image reconstruction. With a phantom experiment, it

has been shown that ASAM has better metal-artifact reduction performance compared with the current methods.

A significant challenge in security imaging is that due to the large geometry and power consumption, low photon statistics are detected. The detected photons suffer high noise and heavy artifacts. Image-domain regularized iterative reconstruction algorithms can reduce the noise but also result in biased reconstruction. A wavelet-domain penalty is introduced which does not bring in bias and can effectively eliminate steaking artifacts. By combining the image-domain and wavelet-domain penalty, the image quality can be further improved.

When the wavelet penalty is used, a concern is that no empirical way, like in the image-domain penalty, is available to determine the penalty weight. Laplace variational automatic relevance determination (Lap-VARD) method is proposed to reconstruct the image and optimal penalty weight choice at the same time.

Chapter 1

Introduction

1.1 Background

1.1.1 X-ray computed tomography systems

X-ray source and collimation

Nearly all medical CT scanners use the rotating anode design and are oil-cooled to reduce heat damage to the anode. Most CT scanners continuously excite the x-ray tube during the data acquisition process. Some CT scanners can be operated in pulse mode, and by switching the tube current, such as between 90 kVp and 140 kVp, a dual-energy scan can be conducted. A major concern is that even with limited usage, the x-ray tube wears out quickly. Regular calibration is needed to ensure accurate measurement.

Different x-ray CT systems require different fan- or cone-beam geometry. Collimation to form the beam shape is accomplished by putting two lead pieces in front of the x-ray source. Between these two lead pieces, there is an adjustable slit to control the thickness of the

beam. In single-slice CT systems, the thickness covers only a single detector slice, and in multi-slice CT systems, a larger slit allows the beam to cover multiple detector slices.

The ideal x-ray CT requires a mono-energetic source. However,conventional x-ray sources generate beams with continuous spectra. The low energy photons are undesirable because they are more likely to be absorbed in the patient's body. They contribute to the patient radiation dose but do not contribute to image formation. To have a better approximation to a mono-energetic source, the lower energy photons are filtered before the x-ray beam enters the patient's body. Figure 1.1 shows the relative intensity of x-ray photons in different propagating stages.

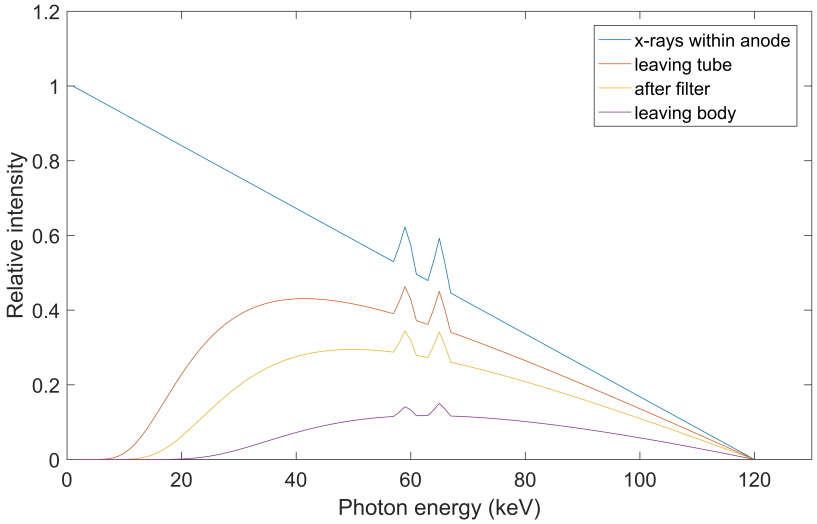


Figure 1.1: Relative intensity of x-ray photons.

CT detector

Most modern x-ray CT systems use solid-state detectors (SSD). The typical design of SSD contains two major components. The first one is a scintillator which interacts with x-ray

photons and generates photoelectrons by the photoelectric effect. These photoelectrons then get de-excited and emit visible light. Typical materials from which scintillators are made include cadmium tungstate, sodium iodide, bismuth germanate, and cesium iodide. The second major component of SSD is the photo-diode which collects the visible light emitted from the scintillator. Photon-diodes convert the light into electric current. The x-ray intensity is presented by the integrated current signal strength. This kind of detector is also called an energy-integrating detector.

Another category of x-ray detector which has recently gotten great attention and development is the energy-resolved detector. Energy-resolved detectors work in a photon-counting mode that keeps track of the energy of every detected photon. A detailed introduction will be presented in Chapter 3.

Gantry

The gantry of a CT system holds the x-ray tube and detectors in opposite directions from the center so that they can be rotated around the patient and acquire the measurement data at the same time. To reduce the patient dose, the gantry should rotate rapidly. A typical scanning time of a 2-D scan is under 1 second. The continuous rotation requires a high standard in mechanical and electrical design.

1.1.2 CT image reconstruction

Data formation

When an x-ray photon with energy E penetrates the scanned object, the survival probability of this x-ray photon follows Beer's law,

$$P(E) = e^{-\mu(E)l}, \tag{1.1}$$

where $\mu(E)$ is the linear attenuation coefficient of the object at energy E, and l is the path length this x-ray photon passes through. If the incident number of photons with energy E is $I_{in}(E)$, then following the survival probability, the expected number of survived photons is

$$I_{out}(E) = I_{in}(E)e^{-\mu(E)l}. \tag{1.2}$$

However, as shown in Figure 1.1, a mono-energetic x-ray source is not practical in CT systems. With the most widely used energy-integrating detectors, the measurement signal measures the photon flux integrated over the whole spectrum. The ideal air-scan normalized transmission signal of the yth source-detector pair is modeled as

$$\frac{I_d(y)}{I_0(y)} = \frac{\int_E I_0(y,E)D(E)e^{-\int_y \mu(r,E)dr}dE}{\int_E I_0(y,E)D(E)dE}, \tag{1.3}$$

where $D(E)$ is the detector-response function, $\mu(r,E)$ is the linear attenuation coefficient at position r, $\int_y \mu(r,E)dr$ is the line integral of the attenuation coefficient along the yth source-detector pair, and $I_0(y)$ is the air-scan intensity, which requires a separate calibration scan.

The transmission model above can be rewritten as

$$\frac{I_d(y)}{I_0(y)} = \int_E \Psi(y,E) e^{-\int_y \mu(r,E)dr} dE, \tag{1.4}$$

where

$$\Psi(y,E) = \frac{I_0(y,E)D(E)}{\int_{E'} I_0(y,E')D(E')dE'} \tag{1.5}$$

is the normalized energy-fluence spectrum.

Analytic image reconstruction

Theoretically, the transmission signal model in (1.4) is correct, but is also mathematically intractable to reconstruct the term $\mu(r,E)$ due to the presence of integration over energy. To simplify the inverse problem, the concept of effective energy $\bar{E}$ is used, such that the transmission model is simplified to

$$\frac{I_d(y)}{I_0(y)} = e^{-\int_r \mu(r,\bar{E})dr} \tag{1.6}$$

Analytic image reconstruction methods to solve the inverse problem, including the most widely used filtered backprojection (FBP) algorithm for 2-D scans and the FDK [1] algorithm for cone-beam scans, can be applied directly on the transmission signal. If the FBP algorithm is used, the reconstructed image is

$$\mu(r,\bar{E}) = FBP\bigg(-\log\Big(\frac{I_d}{I_0}\Big)\bigg). \tag{1.7}$$

The choice of the effective energy $\bar{E}$ is important in the reconstruction. However, as shown in Figure 1.1, the detected photons have higher effective energy because lower energy photons

are more likely to be absorbed in the scanned object. A preprocessing step called beam hardening correction is applied to the transmission data to account for the effective energy mismatch between incident photons and the detected photons.

The FBP reconstruction with beam hardening correction is given by

$$\mu(r, \bar{E}) = FBP\bigg(-\log\Big(BH\Big(\frac{I_d}{I_0}\Big)\Big)\bigg). \tag{1.8}$$

The reconstructed linear attenuation coefficient image μ is $\bar{E}$ related, which means that even the same scanned object will have different $\mu(\bar{E})$ in different CT systems. An even worse case is that the same detector, due to the regular x-ray tube calibration or x-ray tube replacement, may produce different images for the same scanned object. To compare the images from different scanners, the linear attenuation coefficient image is rescaled into CT numbers in Hounsfield units (HU),

$$\mathbf{HU} = 1000\bigg(\frac{\mu(\bar{E})}{\mu_{\text{water}}(\bar{E})} - 1\bigg), \tag{1.9}$$

where $\mu_{\text{water}}(\bar{E})$ is the linear attenuation coefficient of water at effective energy $\bar{E}$. Clearly, HU $= 0$ for water, and HU $= -1000$ for air. In clinical scans, the largest CT numbers can reach about 1000 for bony tissues. In industrial x-ray CT, the CT number can surpass 3000 for heavy metals.

Statistical image reconstruction

Besides the analytic reconstruction algorithm introduced above, statistical image reconstruction (SIR) algorithms have been developed to reconstruct the linear attenuation coefficient

image. The general idea of SIR is to fit the measurement data into a known distribution parameterized by the ground truth and formulate a maximum-likelihood estimation problem. Compared with the analytic algorithms, SIR takes the noise model into consideration. Combined with the prior knowledge as the penalty term, SIR can effectively suppress the noise level and reduce the artifacts.

For energy-integrating detectors, the signal statistics theoretically follow a compound Poisson model. A modified Poisson model is used as an accurate approximation [2]. To minimize the negative Poisson log-likelihood function, the alternating minimization (AM) algorithm [3, 4] is used. The statistical model is that the transmission signals follows the Poisson distribution with mean determined by Beer's law, and its negative log-likelihood is given by

$$L(\mu) = \sum_y [q(y) - I_d(y) \log q(y)], \tag{1.10}$$

where

$$q(y) = I_0(y) e^{-\int_r \mu(r) dr}. \tag{1.11}$$

For computational purposes, $q(y)$ is discretized as

$$q(y) = I_0(y) e^{-\sum_x h(y|x)\mu(x)}, \tag{1.12}$$

where $h(y|x)$ is the CT system matrix elements that represent the path-length contribution of pixel (voxel) x to the source-detector pair y ray path. The operation $\sum_x h(y|x)\mu(x)$ is also called forward-projection. [1]

[1] In this dissertation, the system matrix computation in Chapter 2, 3, and 4 is done with the object-constrained computed tomography (OCCT) software package developed by Dr. David G. Politte at Washington University School of Medicine. The system matrix in Chapter 5 is provided by Dr. Soysal Degirmenci.

Prior knowledge like neighborhood smoothness of images can be added to the negative log-likelihood function as a penalty term,

$$L(\mu) = \sum_{y} [q(y) - I_d(y) \log q(y)] + R(\mu). \tag{1.13}$$

If the penalty term is treated as a prior distribution, a maximum a posteriori (MAP) estimator is obtained by minimizing (1.13).

1.2 Contributions of this work

The main contributions of this dissertation include:

- Designed a DECT phantom experiment and showed that the traditional sinogram-based decomposition method has a systematic error. The systematic error increases as the radiation dose level decreases.
- Proved that the systematic bias in the sinogram-based decomposition method is due to low photon statistics and lack of CT system geometry involvment.
- Proposed a JSIR framework using the DEAM algorithm. The DECT phantom experiment demonstrated that DEAM could effectively eliminate the systematic error in the sinogram-based decomposition method even in an ultra-low-dose scenario.
- Extended the dual-energy result into PCCT systems. Simulation studies showed that PCCT also suffers from biased estimation with the sinogram-based decomposition method.

- Proposed a JSIR framework using the MEAM algorithm for PCCT. Simulation studies showed that the bias could be effectively eliminated.
- Proposed the ASAM algorithm, which can incorporate the advantages of the pre-processing methods and statistical image reconstruction algorithms to reduce the metal artifacts.
- Proposed the wav-AM algorithm for low-dose x-ray CT scans. Explored the feasibility of wavelet-domain regularization, and derived a dual-domain regularized image reconstruction algorithm to generate an unbiased image with low noise and artifacts.
- Proposed a data-driven Lap-VARD algorithm that automatically determines the optimal penalty weight choice for the wavelet-domain penalty.

1.3 Organization of this dissertation

Chapter 2 introduces the DEAM algorithm. The performance of DEAM is shown by a DECT phantom experiment, in which the bias in the sinogram-domain decomposition method is eliminated.

The DECT results are extended to PCCT in Chapter 3, the simulation results show that the biased estimation with sinogram-based decomposition remains in PCCT and can be eliminated by MEAM.

In Chapter 4, the ASAM is introduced to reduce metal artifacts. Through patient phantom simulation and physical phantom experiment, ASAM is found to outperform the currently used metal-artifact reduction methods.

The wav-AM algorithm is proposed in Chapter 5. The wavelet-domain penalty can overcome some challenges in the image-domain regularization. Combining the wavelet-domain penalty and image-domain penalty can further improve the performance.

In Chapter 6, an automatic wavelet-domain penalty weight determination framework, Lap-VARD, is proposed to determine the optimal penalty weight and reconstruct the image at the same time.

Chapter 2

Dual-energy x-ray CT alternating minimization algorithm (DEAM)

2.1 Highlights

In this chapter, we introduce an iterative reconstruction algorithm for dual-energy x-ray CT material decomposition. Compared with the current widely used sinogram-based decomposition method, which has a large bias when the radiation dose level is low, the DEAM generate nearly unbiased estimates.

2.2 Motivation

The use of the effective energy makes the CT image reconstruction problem much easier. However, at the same time, the energy-related information of the linear attenuation coefficient is lost. To accurately reconstruct the energy-related linear attenuation coefficient $\mu(E)$, the idea of spectral CT was proposed. Dual-energy CT is the first implementation

of spectral CT. The energy-related linear attenuation coefficient image can be reconstructed by the sinogram-based decomposition method.

However, when the dose level is low, the widely used sinogram-based decomposition method generates biased reconstruction. To reduce this bias, a joint-statistical image reconstruction framework is proposed.

2.3 Notation

In this section, the notations used in this chapter are summarized.

- d: measured transmission data.
- Q: estimated mean of the transmission data.
- c: basis-material image.
- l: path length of basis material.
- I: air-scan source intensity.
- H: system matrix.
- y: source-detector pair index.
- x: image pixel (voxel) index.
- j: basis-material index.
- k: spectrum index.
- μ: linear attenuation coefficient

2.4 Background

2.4.1 Dual-energy x-ray CT (DECT)

Dual-energy x-ray CT technique scans patients with two different x-ray spectra so that additional information about the tissue properties are provided. Dual-energy scans generate two spectra either by switching the x-ray tube voltage or by running two tubes at different voltages. The settings of 80 kVp and 140 kVp are commonly used because these two spectra provide the maximum difference and least overlap between the spectra. A tube voltage lower than 80 kV is not recommended because the x-ray beam is attenuated by the patient body too much, and sufficient photons cannot be detected. A tube voltage higher than 140 kV is also not used due to hardware limitations. If necessary, extra filters can be used to harden the high energy spectrum and further reduce the overlap between the high and low energy spectra.

Current DECT systems still use energy-integrating detectors which do not provide energy-related information. In the future, detectors with the ability to resolve the energy of each photon, like the cadmium-based semiconductors used in photon-counting CT systems, can be integrated to improve detector performance.

At present, DECT is implemented in various ways. Among them, sequential scanning, dual-source CT, and fast voltage switching are commercially available. Sequential scanning collects the data by performing the single energy scan twice with different tube voltages. The dual-source CT uses two sets of sources with different tube voltage settings and detectors arranged at an angular offset that operate at the same time to acquire the data with different tube voltage settings. The fast voltage switching technique rapidly alternates the tube

voltage between a high value and a low value. The data is collected twice for every projection or at adjacent projections.

Because the tissue attenuation property is energy and material related, the additional energy information provided by DECT enables the estimation of two independent material parameters that are insensitive to the x-ray spectra. This is a major advantage over single energy CT because it uses the "effective energy" which loses a lot of the spectral information.

2.4.2 Basis vector model (BVM)

The BVM assumes that the energy-dependent tissue attenuation coefficient $\mu(x, E)$ can be represented by a linear combination of those of several basis materials,

$$\mu(x, E) = \sum_j c_j(x)\mu_j(E), \tag{2.1}$$

where $c_j(x)$ is the basis-material j component weight at image pixel (voxel) x, and $\mu_j(E)$ is the linear attenuation coefficient (LAC) of material j at energy E. The selection of basis materials has been discussed by Williamson et al. [5] In this chapter, two basis materials, polystyrene and an aqueous calcium chloride solution, are chosen.

2.4.3 Image-based decomposition

In the image-domain decomposition, the DECT measurements are reconstructed separately based on the assumption that the beam hardening correction has been applied to the transmission data of different energy measurements. Single-energy image reconstruction methods

are used to reconstruct high- and low-energy images with their corresponding effective energies.

From the BVM, the LAC image can be decomposed as a linear combination of basis-material constituent images

$$\mu_k(x) = \sum_j \mu_{j,k} c_j(x), \tag{2.2}$$

where $k \in \{L, H\}$ is the high- and low-energy measurement index, $\mu_{j,k}$ is the LAC of basis material j at the corresponding energy k measurement, and $c_j(x)$ is the basis-material j component image at pixel x.

Equation (2.2) can be rewritten in matrix form as

$$\mu(x) = M\mathbf{c}(x), \tag{2.3}$$

where $\mu(x) \in R^k$ is the reconstructed high- and low-energy LAC vector at pixel x, M is the mixing matrix with the matrix element $\mu_{j,k}$, and $\mathbf{c}(x) \in R^j$ is the basis-material component vector at pixel x.

The basis-material component vector can be directly recovered by

$$\mathbf{c}(x) = M^{-1}\mu(x). \tag{2.4}$$

From (2.4), the basis-material component image c relies on accurate single image reconstruction results $\mu(x)$. A major problem in the image-domain sinogram decomposition method is that even though the beam hardening correction can be applied to high- and low-energy measurements, the beam hardening effect is not fully compensated.

2.4.4 Sinogram-based decomposition method

The sinogram-domain decomposition method [6, 7] is a two-step approach to reconstruct the basis-material images. In the first step, the basis-material sinograms are decomposed from the dual-energy measurements. In the second step, the corresponding basis-material images are generated via the FBP algorithm from the decomposed sinograms.

Let $\psi_k(y, E), k \in \{L, H\}$ be the detector response-weighted spectrum of the low- and high-energy scans for photon energy E and source-detector pair y, where E is discretized at 1 keV intervals. Let $I_{0,k}(y)$ be the corresponding air-scan detector response. For a scanned object, the expected transmission sinogram can be modeled as

$$Q_k(y) = I_{0,k}(y) \sum_E \psi_k(y, E) e^{-\sum_x h(y|x)\mu(x,E)}, \tag{2.5}$$

where $\mu(x, E)$ is the photon linear attenuation coefficient of energy E at image pixel x and $h(y|x)$ is the dual-energy CT system matrix element which represents the effective length of the intersection between the ray path for source-detector pair y and image pixel x. Based on the BVM, the sinograms of the two basis materials are defined as

$$l_j(y) = \sum_x h(y|x) c_j(x). \tag{2.6}$$

Although it has been proven that the signal statistics of energy-integrating detectors theoretically follow a compound Poisson model, our prior work shows that for clinical CT, a modified Poisson model is an accurate approximation [8, 2]. With this approximation, we assume the measurement $d_k(y)$ follows a Poisson distribution with mean equal to $Q_k(y)$.

With (2.6), equation (2.5) can be rewritten as

$$Q_k(y) = I_{0,k}(y) \sum_E \psi_k(y, E) e^{-\sum_j l_j(y)\mu_j(E)}. \tag{2.7}$$

Based on the Poisson noise model, the maximum log-likelihood estimation (MLE) problem with respect to $l_j(y)$ is defined as

$$\min_{l_j(y)} \sum_k [Q_k(y) - d_k(y) \log Q_k(y)]. \tag{2.8}$$

By solving this MLE problem for each source-detector pair y independently, two basis-material sinograms are generated. After the decomposition, the two corresponding images of the basis-material weights $c_j(x)$ are obtained using the FBP algorithm. A clear advantage of the sinogram-based decomposition method over the image-domain decomposition method is that the beam hardening effect is incorporated into the statistical model in (2.5), and no beam hardening correction is needed.

2.5 Dual-energy joint statistical image reconstruction

In the previous section, the sinogram-based material decomposition method was introduced. The basis-material images are reconstructed from the basis-material sinograms. It has been shown by Fu et al. [9] that using the transmission pre-log data has an advantage over using the post-log attenuation sinogram especially when the dose level is low. In this section, instead of using the two-step sinogram-based decomposition method, a joint statistical image reconstruction (JSIR) framework using the dual-energy alternating minimization algorithm

is proposed to reconstruct the basis-material images directly from the low- and high-energy transmission data.

2.5.1 Statistical model

In order to directly reconstruct the basis-material image c, with BVM, the line-integral of LAC for source-detector pair y is rewritten as

$$\mu(y, E) = \sum_j \mu_j(E) \sum_x h(y|x) c_j(x). \tag{2.9}$$

Due to the presence of $h(y|x)$, the assumption that each of the basis-material sinogram pixels $l_j(y)$ is independent is not valid anymore. The maximum log-likelihood estimation (MLE) problem becomes

$$\min_c \sum_k \sum_y [Q_k(y, c) - d_k(y) \log Q_k(y, c)], \tag{2.10}$$

where

$$Q_k(y, c) = I_{0,k}(y) \sum_E \psi_k(y, E) e^{-\sum_j \mu_j(E) \sum_x h(y|x) c_j(x)}. \tag{2.11}$$

2.5.2 Alternating minimization

Instead of solving problem (2.10), the following problem is solved

$$\begin{aligned}\min_{p,q} F(p,q) &= \sum_y \sum_k \sum_E I(p_k(y,E)||q_k(y,E)) \\ &= \sum_y \sum_k \sum_E p_k(y,E)\log\frac{p_k(y,E)}{q_k(y,E)} - p_k(y,E) + q_k(y,E) \\ \text{s.t.} \quad &\sum_E p_k(y,E) = d_k(y) \\ & q_k(y,E) = I_{0,k}(y)\psi_k(y,E)e^{-\sum_j \mu_j(E)\sum_x h(y|x)c_j(x)}.\end{aligned} \tag{2.12}$$

Theorem 2.1. *Solving problem* (2.12) *is equivalent to solving problem* (2.10).

The proof of Theorem 2.1 is in Appendix A.2.

To minimize the cost function with respect to p, only the terms that contain p are considered

$$\min_p F_p(p) = \sum_{y,k,E}\Big[p_k(y,E)\log\frac{p_k(y,E)}{q_k(y,E)} - p_k(y,E) + q_k(y,E)\Big], \tag{2.13}$$

$$\text{s.t.} \sum_E p_k(y,E) = d_k(y).$$

After solving the Lagrangian multiplier, the minimizer of problem (2.13) is

$$p_k(y,E) = \frac{q_k(y,E)}{\sum_E q_k(y,E)} d_k(y). \tag{2.14}$$

From the definition of q in 2.12, q is a function of c. Minimizing the cost function with respect to q is equivalent to minimizing the cost function with respect to c. After solving p when q is kept fixed, only the terms that contain c are considered to minimize the cost

function with fixed p,

$$\begin{aligned}\min_c F_c(c) &= \sum_{y,k,E} -p_k(y,E)\log q_k(y,E) + q_k(y,E) \\ &= \sum_{y,k,E}\Big[I_{0,k}(y)\psi_k(y,E)e^{-\sum_j \mu_j(E)\sum_x h(y|x)c_j(x)} + p_k(y,E)\sum_j \mu_j(E)\sum_x h(y|x)c_j(x)\Big].\end{aligned} \tag{2.15}$$

In (2.15), the values of $c_j(x)$ are coupled with each other. The direct computation of the gradient with respect to c would require a huge amount of computation, which is not feasible. Instead of directly computing the gradient of the cost function, the covex decomposition lemma [10]

Lemma 2.1. *Let $f(x)$ be a convex function, then*

$$\begin{aligned} f(\sum_i x_i) &\le \sum_i r_i f(\frac{1}{r_i}x_i) \\ s.t.\ \sum_i r_i &= 1,\ r_i > 0. \end{aligned} \tag{2.16}$$

is used to decouple c. Based on the fact that $f(x) = e^{-x}$ is a convex function, a surrogate function of the original cost function will be generated, and it is of the form

$$\hat{F}_c(c) = \sum_{y,k,E}\Big[\hat{q}_k(y,E)\sum_{j,x}\frac{\mu_j(E)h(y|x)}{Z_j(x)}e^{-Z_j(x)\left(c_j(x)-\hat{c}_j(x)\right)} + p_k(y,E)\sum_j \mu_j(E)\sum_x h(y|x)c_j(x)\Big], \tag{2.17}$$

where

$$\hat{q}_k(y,E) = I_k(y,E)e^{-\sum_{j,x}\mu_j(x)h(y|x)\hat{c}_j(x)}, \tag{2.18}$$

$$Z_j(x) = \sum_{y,E}\mu_j(E)h(y|x). \tag{2.19}$$

$\hat{c}_j(x)$ is the current estimate of $c_j(x)$ and Z is a parameter that guarantees convergence. The surrogate function's derivative with respect to $c_j(x)$ is

$$\frac{\partial \hat{F}_c}{c_j(x)} = \sum_{y,k,E} \Big[-\hat{q}_k(y,E)\mu_j(E)h(y|x)e^{-Z_j(x)\left(c_j(x)-\hat{c}_j(x)\right)} + p_k(y,E)\mu_j(E)h(y|x) \Big]. \tag{2.20}$$

Setting this derivative equal to zero gives

$$e^{-Z_j(x)\left(c_j(x)-\hat{c}_j(x)\right)} = \frac{\sum_{y,k,E} p_k(y,E)\mu_j(E)h(y|x)}{\sum_{y,k,E} q_k(y,E)\mu_j(E)h(y|x)}. \tag{2.21}$$

The final $c_j(x)$ update equation is

$$c_j(x) = \hat{c}_j(x) - \frac{1}{Z_j(x)} \log \left(\frac{\sum_{y,k,E} p_k(y,E)\mu_j(E)h(y|x)}{\sum_{y,k,E} q_k(y,E)\mu_j(E)h(y|x)} \right). \tag{2.22}$$

Equation (2.22) gives a closed-form update for the basis-material images. However, when the scanning dose level is low, the image quality suffers heavy noise. To suppress the noise, a neighborhood smoothing penalty is added to the surrogate function. In this chapter, a Huber-type penalty is selected with the form [11, 12]

$$R(c) = \sum_j \sum_x \sum_{\tilde{x}\in\mathcal{N}_x} w_{x\tilde{x}} \phi\big(c_j(x) - c_j(\tilde{x})\big), \tag{2.23}$$

where

$$\phi(t) = \frac{1}{\delta^2}\big(\delta|t| - \log(1+\delta|t|)\big). \tag{2.24}$$

$\mathcal{N}_x$ is the set of neighborhood of pixel x, and the corresponding weights $w_{x\tilde{x}}$ are equal to the inverse distance between pixels x and $\tilde{x}$. One advantage of this penalty is that when t is close to 0, $\phi(t)$ performs like the quadratic function $|t|^2$, and when t is far away from 0, $\phi(t)$ performs like the total variation penalty $|t|$ [13]. The parameter δ controls the transition

point between the quadratic region and the linear region. Another advantage of this Huber-type penalty is that it is twice continuously differentiable at 0, while the total variation penalty does not have a continuous second order derivative.

The regularized DEAM cost function is

$$\min_c \hat{F}_c(c) = \sum_{y,k,E} \Bigg[\hat{q}_k(y,E) \sum_{j,x} \frac{\mu_j(E) h(y|x)}{Z_j(x)} e^{-Z_j(x)\left(c_j(x) - \hat{c}_j(x)\right)} + p_k(y,E) \sum_j \mu_j(E) \sum_x h(y|x) c_j(x) \Bigg] + \lambda R(c). \tag{2.25}$$

The basis-material image pixels $c_j(x)$ are coupled in the penalty term $\phi(c_j(x) - c_j(\tilde{x}))$. Again, the convex decomposition lemma is used to decouple them. The resulting cost function is

$$\min_c \hat{F}_c(c) = \sum_{y,k,E} \Bigg[\hat{q}_k(y,E) \sum_{j,x} \frac{\mu_j(E) h(y|x)}{Z_j(x)} e^{-Z_j(x)\left(c_j(x) - \hat{c}_j(x)\right)} + p_k(y,E) \sum_j \mu_j(E) \sum_x h(y|x) c_j(x) \Bigg] + \lambda \sum_j \sum_x \sum_{\tilde{x} \in \mathcal{N}_x} \phi\big(2c_j(x) - \hat{c}_j(x) - \hat{c}_j(\tilde{x})\big). \tag{2.26}$$

The corresponding first and second order derivatives of the decoupled cost function are

$$\frac{\partial \hat{F}_c}{\partial c_j(x)} = \sum_{y,k,E} \Big[- \hat{q}_k(y,E) \mu_j(E) h(y|x) e^{-Z_j(x)\left(c_j(x) - \hat{c}_j(x)\right)} + p_k(y,E) \mu_j(E) h(y|x) \Big] + \lambda \sum_{\tilde{x} \in \mathcal{N}_x} w_{x\tilde{x}} \frac{2c_j(x) - \hat{c}_j(x) - \hat{c}_j(\tilde{x})}{1 + \delta |2c_j(x) - \hat{c}_j(x) - \hat{c}_j(\tilde{x})|} \tag{2.27}$$

and

$$\frac{\partial \hat{F}_c^2}{\partial^2 c_j(x)} = \sum_{y,k,E} \hat{q}_k(y,E)\mu_j(E)h(y|x)Z_j(x)e^{-Z_j(x)\left(c_j(x)-\hat{c}_j(x)\right)} + \lambda \sum_{\tilde{x}\in\mathcal{N}_x} w_{x\tilde{x}} \frac{2}{\left(1+\delta|2c_j(x)-\hat{c}_j(x)-\hat{c}_j(\tilde{x})|\right)^2}, \tag{2.28}$$

The pseudo-code of the DEAM algorithm is presented in Algorithm 1.

Algorithm 1: Dual-energy alternating minimization algorithm

Initialize $c_j(x)$, and compute the corresponding $\hat{q}_k(y, E)$ and $\hat{p}_k(y, E)$.
if without penalty **then**
  for $n = 0$ to $N - 1$ **do**
    foreach j **do**
      Set $c_j(x)^{(n)}$ as the current estimate.
      Update $\hat{q}_k(y, E)$ with (2.18).
      Update $\hat{p}_k(y, E)$ with (2.14).
      Update $c_j(x)^{(n+1)}$ with (2.22).
    end
  end
end
if with penalty **then**
  for $n = 0$ to $N - 1$ **do**
    foreach j **do**
      Set $c_j(x)^{(n)}$ as the current estimate.
      Update $c_j(x)^{(n+1)}$ with Newton's method, and the first order and second order derivatives are from (2.27) and (2.28).
    end
  end
end

2.6 Physical phantom experiment

2.6.1 Experimental setup

A customized phantom was used to test the performance of the DEAM algorithm. The phantom is water based and has a cylindrical shell with a diameter of 215 mm as shown in Figure 2.1. Liquid samples were contained in plastic bottles with a diameter of 31 mm and they were mounted to an out-of-field platform inside the phantom. To get an accurate reference linear attenuation coefficient for quantitative assessment, four pure liquid materials, water, propanol, ethanol, and butanol were selected.

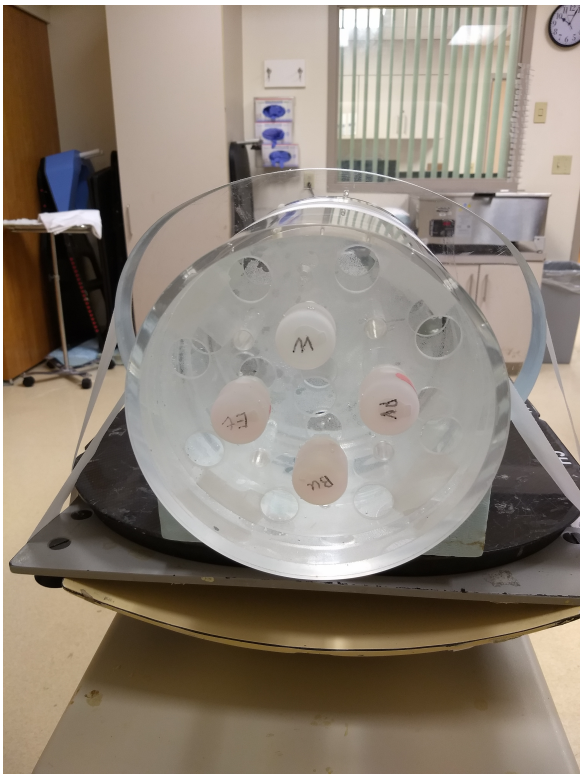


Figure 2.1: Image of the test phantom used in the experiment (back view). The phantom is water based and contains 4 sample inserts.

The test phantom was scanned on a Philips Brilliance Big Bore CT scanner (Philips Medical System, Cleveland, OH) using an axial protocol with 0.75 mm $\times$ 4 collimation to minimize the influence of scatter. The low energy scan used a spectrum with 90 kVp, and the high energy scan used a spectrum with 140 kVp. Raw transmission sinograms, with the vendor's beam-hardening corrections disabled, were exported. To test the accuracy with respect to the dose level, the (90/140 kVp) scans were performed at four dose levels: high (400 mAs/200 mAs), medium (200 mAs/100 mAs), low (50 mAs/25 mAs), and ultra-low (15 mAs/mAs). The central axis spectra of these two scans were determined using the well-validated equivalent spectrum method [14, 15, 16] where the detector-response weighted spectra were determined by fitting the Birch-Marshall model [17] to the measurement beam

transmission profiles through stacks of high purity aluminum and copper filters with varying thickness. Details of the experimental setup were described by Evans et al. [16].

In both the sinogram-basd decomposition method and the DEAM algorithm, the reconstructed image has a pixel size of 0.98 mm $\times$ 0.98 mm and a slice thickness of 3 mm. The raw sinograms were exported from the scanner and were preprocessed by the tool provided by the vendor without beam hardening correction. In the DEAM algorithm, the parameters were set to be $\delta = 100$ and $\lambda = 5$ for a trade-off between noise reduction and resolution which is slightly higher than the scanner's resolution for typical soft-tissue contrast. The background events were assumed to be zero for this narrow-beam collimation protocol.

2.6.2 Performance assessment

The basis-material images $c_j(x)$ were reconstructed with both the sinogram-based decomposition method and the DEAM algorithm. In this experiment, polystyrene and 23% $CaCl_2$ aqueous solution were used as the basis materials, which have been proven to parameterize the LAC to within 1-2% accuracy [5, 18, 19]. The LAC were computed based on the BVM model according to

$$\mu(x, E) = \mu_1(E)c_1(x) + \mu_2(E)c_2(x), \tag{2.29}$$

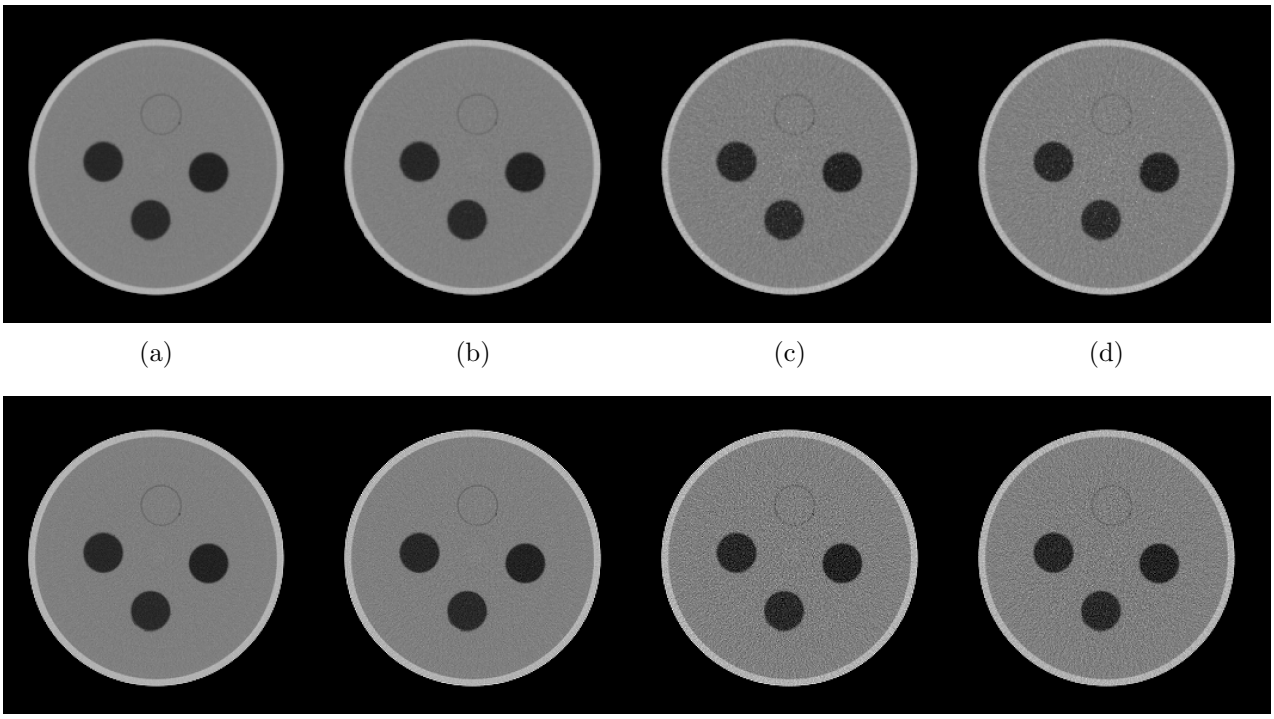
where c_1 and c_2 are the basis-material images of polystyrene and $CaCl_2$ aqueous solution, and μ_1 and μ_2 are the corresponding LAC. The reconstructed LAC were compared with the reference ground-truth values. The ground-truth values of water and other sample materials were from the National Institute of Standards and Technology (NIST) XCOM database [20].

To better show the results quantitatively, the CT numbers were computed based on

$$\mathrm{HU}(E) = 1000\left(\frac{\mu(E)}{\mu_{\mathrm{water}}(E)} - 1\right). \tag{2.30}$$

2.6.3 Results

The reconstructed CT number images using both the sinogram-based decomposition method and the DEAM algorithm are shown in Figure 2.2. All CT number images were computed



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

(e) (f) (g) (h)

Figure 2.2: (a) - (d) show the reconstructed CT number images with the sinogram-based deomposition method at 60 keV and at different dose-level settings: (a) 400 mAs/200 mAs, (b) 200 mAs/100 mAs, (c) 50 mAs/25 mAs, and (d) 15 mAs/15 mAs for the 90 kVp and 140 kVp scans respectively. (e) - (h) show the reconstructed CT number images with the JSIR method at the corresponding dose-level settings.

at 60 keV, which is approximately the effective energy of clinical scans. With decreased dose level, a direct observation is that the noise increases with both methods. In this experiment, all the transmission sinograms were normalized with respect to the air scan, and the same parameters λ and δ were used in the DEAM algorithm. To further reduce the noise level, a large λ is required at the cost of loss of resolution. In our previous study, it was demonstrated that compared with the sinogram-based decomposition method, the DEAM algorithm has superior performance in noise reduction and artifact suppression [19].

To assess the quantitative accuracy of both methods, four homogeneous regions in the sample inserts were selected which are shown in Figure 2.3. The average LAC of these four regions

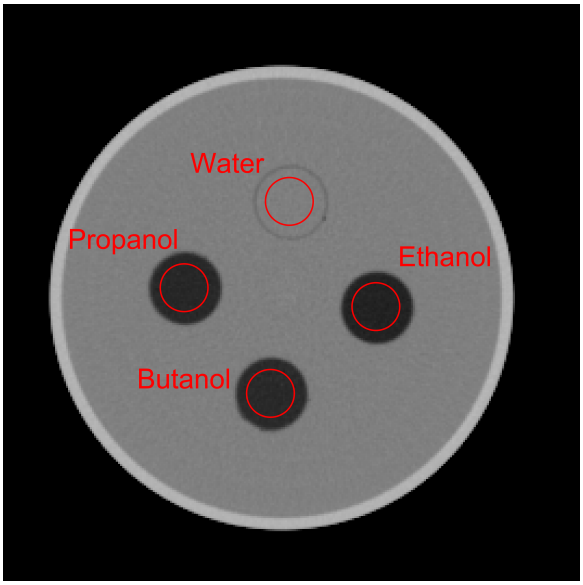


Figure 2.3: Four homogeneous regions selected inside the sample inserts.

were computed in forms of CT numbers over a spectral range from 30 keV to 140 keV. We compared the CT numbers with the reference values from the NIST XCOM database, and the relative errors are shown in Figure 2.4. In high-dose and medium-dose levels, the sinogram-based decomposition method and the DEAM algorithm have very similar accuracy performance, and the average relative errors are close to 0 which means both methods are

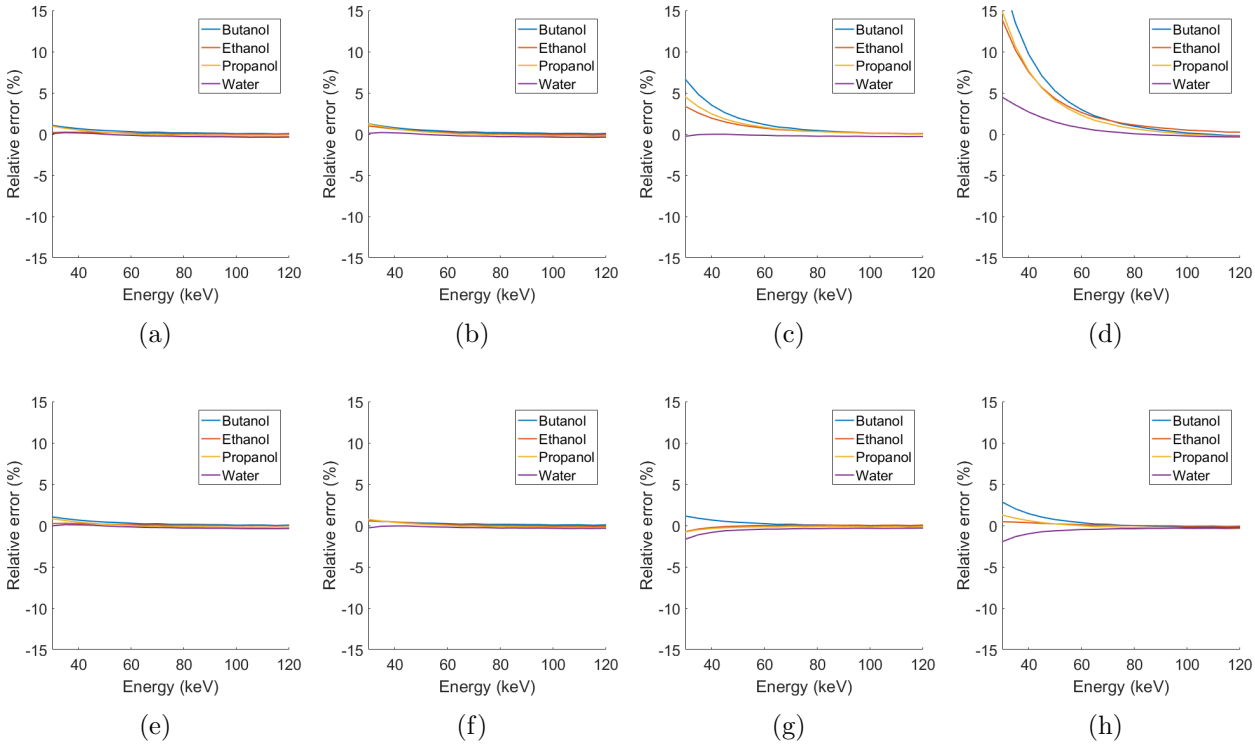


Figure 2.4: (a) - (d) show the relative error in CT numbers using the sinogram-based decomposition method at 60 keV and at different dose-level settings: (a) 400 mAs/200 mAs, (b) 200 mAs /100 mAs, (c) 50 mAs /25 mAs, and (d) 15 mAs / 15 mAs for the 90 kVp and 140 kVp scans respectively. (e) - (H) show the relative error using the JSIR method at the corresponding dose-level settings. The viewing windows are [-300, 300] HU.

nearly unbiased estimators. When the dose level further decreases to the medium-dose setting, 90 kVp with 50 mAs and 140 kVp with 25 mAs, the relative error of the butanol, ethanol, and propanol show significant positive increase especially in the lower energy range. However, the DEAM algorithm shows that all the relative errors of these samples are within $\pm 1\%$ even in the lower energy range. In the most extreme ultra-low dose case, 90 kVp with 15 mAs and 140 kVp with 15 mAs, the sinogram-based method generates huge positive bias which can be as high as 15% in the lower energy range. Using the DEAM algorithm, the relative error of the samples show no significant biases and the maximum values are kept within $\pm 2.5\%$. This experiment shows that the sinogram-based decomposition method is a

biased estimator, and with a decrease in the dose level, the sinogram-based decomposition method has rapidly increased positive bias. The proposed DEAM algorithm, however, is a nearly unbiased estimator, and even in the ultra-low dose dual-energy scans, the DEAM algorithm can still generate very accurate results.

2.6.4 Discussion

Low photon statistics

In the physical phantom experiment, it was shown that the sinogram-based decomposition method is a biased estimator, and the bias is determined by the dose levels. In the sinogram-based decomposition method, the basis-material images are reconstructed from the basis-material sinograms with the FBP algorithm. Because the FBP algorithm is a linear operator, the biases in the images correspond to the biases in the basis-material sinograms. For a single source-detector pair, we assume that the scanned object is a uniform $CaCl_2$ aqueous solution with fixed length l_0. Assuming a mono-energetic x-ray source with energy E is used, the ground truth of the line integrated LAC is $\mu(E) = \mu_{CaCl_2}(E)l_0$. It is proven in Appendix A.1 that the estimated value of the reconstructed LAC has a positive bias, which is $\mathrm{E}(\hat{\mu}(E)) - \mu(E) \geq 0$. Also from Appendix A.1, the sinogram-based decomposition method generates biased estimates and the bias decreases when the dose level increases.

A Monte-Carlo simulation was used to show the relationship between bias and the dose level. The decomposed sinograms have a systematic bias which leads to the bias in the basis-material image [21]. A pair of 90 kVp and 140 kVp detector-response weighted spectra were used. The basis materials were chosen to be polystyrene and $CaCl_2$ aqueous solution. The scanned object was 250 mm of polystyrene. With 1 million Monte-Carlo trials for each

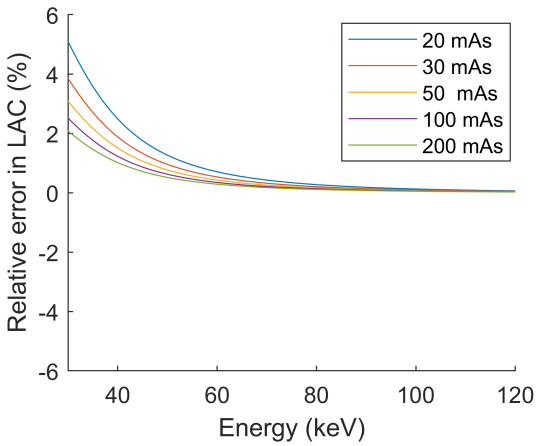
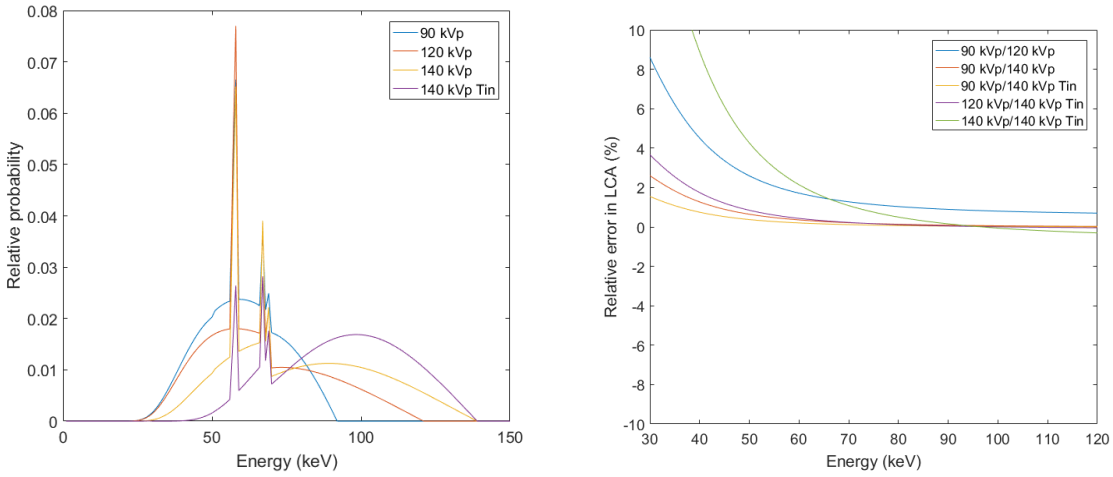


Figure 2.5: Monte-Carlo simulation of the relative estimation error with respect to different dose levels.

dose level from 20 to 200 mAs and with a 90 kVp spectrum (with the dose level of 140 kVp spectrum set computed accordingly), the relative estimation error of the LAC is shown in Figure2.5. With the increased dose level, the relative error monotonically decreases.

Beam hardening - photon starvation at lower energies

Another observation in the physical phantom experiment is that bias is also energy dependent. The bias is more prominent in the lower energy range and decreases with increased energy. This phenomenon is due to the beam hardening effect [19]. Lower energy photons are more likely to be absorbed in the phantom. The result is that the detectors detect very few low-energy photons, and the spectrum at the detectors shifts towards higher energies. Based on the discussion in section 2.6.4, this photon starvation at lower energies results in more substantial bias at lower energies. To compensate the photon starvation at low energies, while keeping the total dose level constant, a spectrum with lower effective energy



(a) (b)

Figure 2.6: Monte-Carlo simulation of different spectra: (a) shows four detector response-weighted spectra; (b) shows the relative error with five sets of dual-energy scans.

should be used. However, this spectrum will cause lower total photon counts due to the beam hardening effect, which will increase the bias.

A Monte-Carlo simulation was used to demonstrate the impact of the spectrum and the beam hardening effect. The scanned object was still 250 mm polystyrene. Four detector response-weighted spectra were employed. The normalized spectra of 90 kVp, 120 kVp, 140 kVp and 140 kVp after 0.25 mm tin filtration with effective energies of 60 keV, 66 keV, 78 keV, and 93 keV are shown in Figure 2.6(a). Five sets of dual-energy scans with the same total photon intensities were simulated: 90 kVp/120 kVp, 90 kVp/140 kVp, 90 kVp/140 kVp with the tin filer, 120 kVp/140 kVp with tin filter, and 140 kVp/140 kVp with tin filter. The average relative errors of the five scans are shown in Figure 2.6(b). To maximize the number of detected photons, the pair of highest effective energy 140 kVp/140 kVp with the tin filter was used. This pair has the largest relative error in low energy range due to the extreme photon starvation from beam hardening effect in this energy range. To compensate

for the low photon statistics at the lower energy range, the pair of 90 kVp/120 kVp was used. Compared with the pair of 140 kVp/140 kVp with the tin filter, the bias in the low energy range of below 50 keV is reduced. However, due to having the overall lowest number of detected photons, the bias in the higher energy range is significant. The other three pairs all have good bias reduction, and a direct observation is that the relative bias decreases when the effective energy gap between the two spectra increases. In this simulation, the pair 90 kVp/140 kVp with tin filter, which has the largest effective energy gap of 33 keV, has the best performance.

Though it has been shown that the effective energy gap should be as large as possible, the low energy spectrum requires a very high tube current to get sufficient transmission data, which will wear out tubes rapidly [22]. There is no need to go lower than 90 kVp.

Effect of system geometry

In the sinogram-based decomposition method, the sinograms are decomposed by solving problem (2.8) independently for each source-detector pair y. The underlying assumption is that the measurement of each basis-material line integral is independent. However, this assumption is not valid. Within a fixed scanning angle (projection slice), we can assume all the projections are independent because there is no overlap between these projections. However, when the gantry rotates to another scanning angle, the projection slice is correlated with the previous one. The sinogram pixels are independent within a single projection slice but are correlated between different projection slices because their projection ray path may intersect and share a common image pixel. This can be explained by the forward-projection

model

$$l_j(y) = \sum_x h(y|x)c_j(x). \tag{2.31}$$

In the DEAM algorithm, the forward-projection model is involved in the statistical model. If we take the forward-projection model as a constraint, the minimization problem of (2.10) is equivalent to the following minimization problem,

$$\min_c \sum_k \sum_y [Q_k(y,c) - d_k(y) \log Q_k(y,c)], \tag{2.32}$$

$$\text{s.t.} \quad l_j(y) = \sum_x h(y|x)c_j(x).$$

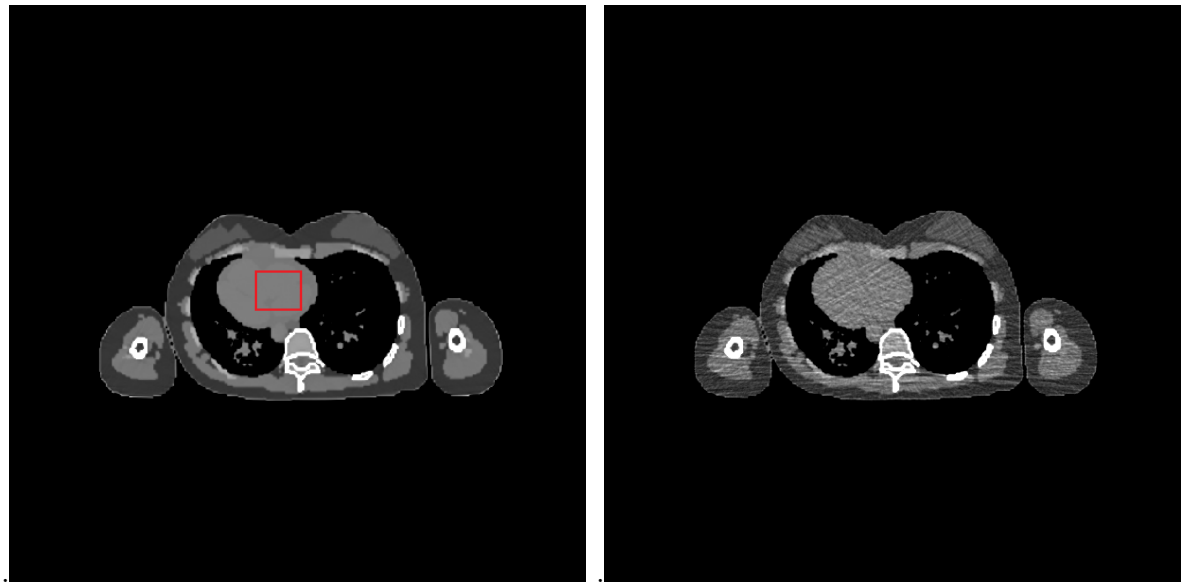
Problem (2.32) is the constrained version of problem (2.8) and can be solved by

$$\begin{aligned} \min_{p,q} F(p,q) = &\sum_y \sum_k \sum_E I(p_k(y,E)||q_k(y,E)) + \beta \sum_j I\big(l_j(y)|| \sum_x h(y|x)c_j(x)\big) \qquad (2.33) \\ \text{s.t.} \quad &\sum_E p_k(y,E) = d_k(y) \\ &q_k(y,E) = I_{0,k}(y)\psi_k(y,E)e^{-\sum_j \mu_j(E) l_j(y)}, \end{aligned}$$

with the line integral alternating minimization algorithm (LIAM) [23]. The parameter β penalizes the discrepancy between the estimated line integrals and the forward projections of the component images. When $\beta = 0$, the forward-projection model (2.6) is completely ignored, and problem (2.33) is equivalent to the sinogram-based decomposition method. When $\beta \to +\infty$, the discrepancy between the line-integral $l_j(y)$ and the forward-projection $\sum_x h(y|x)c_j(x)$ must go to zero, and problem (2.33) is equivalent to the original DEAM problem (2.12). By the choice of β, LIAM provides an intermediate state between the sinogram-based decomposition method and DEAM, and β controls the influence of the

forward-projection model. A phantom simulation was performed to assess the relationship between bias and parameter β.

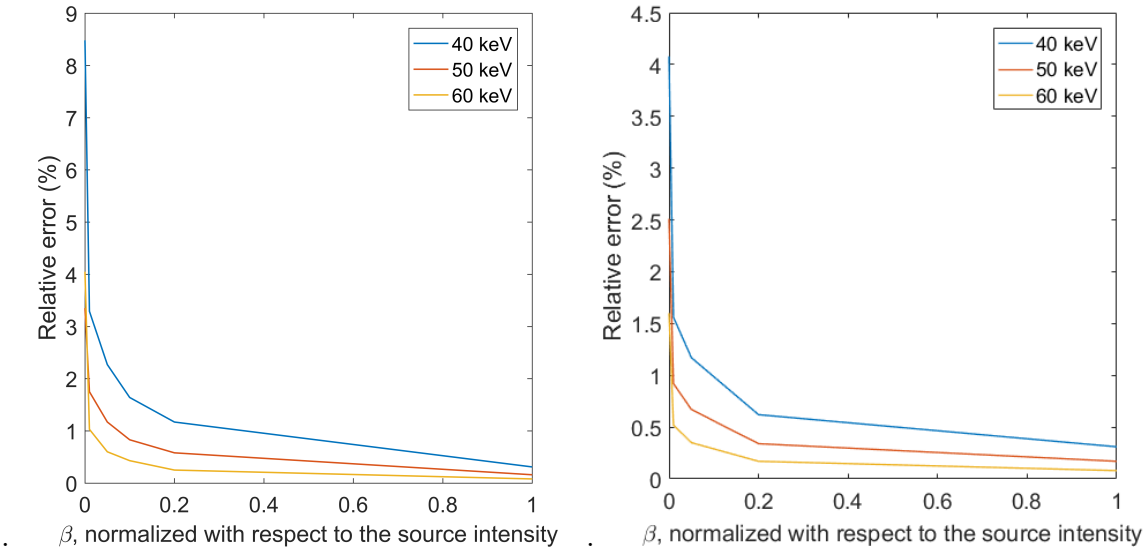
In this simulation, a thorax slice of the International Commission on Radiological Protection (ICRP) phantom [24] was scanned. The system geometry of a Philips Brilliance Big Bore CT scanner was used. The low-energy scan used a spectrum of 90 kVp, and the high energy scan used a spectrum of 140 kVp. Two low-dose levels, 40 mAs/40 mAs and 20 mAs/20 mAs, were simulated. In this simulation, polystyrene and 23% $CaCl_2$ aqueous solution were used



(a) (b)

Figure 2.7: (a) The ground truth CT number image at 50 keV; (b) LIAM reconstructed CT number image at 50 keV of the 40 mAs/40 mAs scan with parameter $\beta = 1$. The viewing windows are [-200, 400] HU.

as the basis materials. The basis-material component images were reconstructed with LIAM. Based on the BVM, the LAC images were computed and then presented in the form of CT numbers. To quantitatively track the impact of the parameter β in LIAM, a homogeneous region was selected to compute the relative error as shown in Figure 2.7(a). In Figure 2.8,



(a) 20 mAs/ 20 mAs scan

(b) 40 mAs/ 40 mAs scan

Figure 2.8: Impact of β on relative error.

when $\beta = 0$, which is equivalent to the sinogram-based decomposition method, the relative error can be as high as 8.5% for the 20 mAs/20 mAs scan and 4% for the 40 mAs/40 mAs scan. With the increase of β, which means a smaller discrepancy between the basis-material line integral and the forward projection of the basis-material component image is allowed, the relative error dramatically decreases. When β is large enough, which is an approximation to DEAM, the relative error is close to zero. The lack of including the effects of the system geometry is the major factor causing the bias in the sinogram-based decomposition method. With DEAM, which directly reconstructs the basis-material component images, the bias can be effectively eliminated.

2.7 Summary

Dual-energy CT systems can extract the energy-related information by doing material decomposition. The conventional sinogram-based decomposition has a significant bias when the dose level is low. The reason for this bias is the photon starvation and lack of system geometry involvement. The DEAM algorithm modifies the statistical model by involving the forward projection into the statistical model and increasing the photon statistics. In a physical phantom experiment, the DEAM algorithm was shown to be a nearly unbiased estimator which can effectively eliminate the systematic bias in the sinogram-based decomposition method even in an ultra-low-dose scenario.

Chapter 3

Multi-energy alternating minimization algorithm (MEAM) for photon-counting CT (PCCT)

3.1 Highlights

In this chapter, based on the DEAM algorithm in the previous chapter, the multi-energy alternating minimization algorithm for photon-counting CT system is introduced to further explore the available energy information.

3.2 Motivation

In the previous chapter, it was shown that the energy information in the x-ray CT scan can be extracted by the usage of a dual-energy scan. In the past decade, another x-ray CT modality, PCCT, has been greatly developed. Compared with conventional CT systems, the

PCCT systems can provide additional energy dependent information which enables material decomposition [25], reduces radiation dose [26], reduces beam hardening effects, and gives higher signal-to-noise ratio [27].

The most common PCCT material decomposition method separates the steps of material decomposition and image reconstruction. In sinogram-based decomposition [28], material separation is performed prior to the basis-material reconstruction. In the image-based decomposition [29, 30], images from different energy bin measurements are reconstructed first, and then the decomposition is applied. Separating these steps has the flaw that the geometry and statistical information are not fully used as stated in Chapter 2. With decreased radiation dose, the material decomposition has dramatically increased bias.

In this chapter, we present a joint image statistical framework using the multi-energy alternating minimization algorithm (MEAM) for PCCT systems. The basis material images are reconstructed directly from the spectral measurements. In a patient phantom simulation, the MEAM algorithm effectively eliminates the bias caused by photon starvation.

3.3 Notation

In this section, the notations used in this chapter are summarized.

- d: measured transmission data.
- Q: estimated mean of the transmission data.
- μ: linear attenuation coefficient
- c: basis-material image.

- l: path length of basis material.
- I: air-scan source intensity.
- H: system matrix.
- y: source-detector pair index.
- x: image pixel (voxel) index.
- j: basis material index.
- k: energy bin index.

3.4 Background

3.4.1 Photon-counting detectors (PCDs)

Most clinical systems use conventional energy-integrating detectors which collect the intensity of x-ray beams, but the energy-integrating detector lacks the ability to resolve the energy spectrum of an x-ray beam, which gives extra information about the scanned object. With energy-integrating detectors, the widely used dual-energy CT systems [31, 32, 33] can track the energy information by using a combination of high/low energy spectrum scans. PCDs with the ability of energy discrimination can collect the x-ray photons in different energy bins simultaneously. With PCDs, the PCCT system can extract the energy information from a single x-ray spectrum, which reduces the radiation dose compared with the dual-energy CT scan. The advantages of PCCT also include reduced noise [34], improved contrast [35], and better tissue identification [36].

However, the measurement data of the PCDs are affected by several major factors that degrade the measurement. The first one is called the *pulse pileup effect* [35, 37], which happens when multiple photons are detected by the detector within a short period of time. The electrical pulse of these photons overlap each other and result in loss of counts and distorted energy measurements. With the development of PCDs, the time resolution has been improved, but due to physical limitations, this distortion cannot be eliminated. Two major correction methods have been proposed to deal with this *pulse pileup effect* [38, 39]. The first one is to calibrate the measurements to compensate for the spectral distortion which requires an extensive amount of measurements [40, 38, 41]. The other one is the model-based compensation method which uses a small number of parameters to model the spectral distortion due to the pileup effect [42].

The second major factor in distorting the measurement is that when an incident photon interacts with the detector, one or multiple electrical charges can be generated through the photoelectric effect, Compton scatter, and the reabsorption of K-escape photons [42]. If the electrical charge is generated near the detector pixel boundaries, the charge will be split and detected by multiple detector pixels at lower energies. These factors can be summarized in a probability distribution for an x-ray photon of energy E_0 to be detected by the detector at energy E. This distribution is called the *spectral response function* (SRF) [42, 43]. To increase the photon statistics, the energy spectral measurement is always binned into several energy bins. By summing the SRF within every single energy bin, the corresponding energy bin detection probability distribution, *bin response function* (BRF), is obtained.

3.4.2 Sinogram-based decomposition

Statistical model

The attenuation properties of matter for x-ray CT are material and energy dependent. Compared with traditional x-ray CT systems, the additional energy information provided by PCCT enables the derivation of material information (i.e., material decomposition). Let $d_k(y), k \in 1...K$ denote the transmission measurement of energy bin k at the yth source-detector pair. Following Lange and Carson [44], the data are modeled as independent Poisson random variables with the probability distribution

$$Pr[d_k(y) = n] = \frac{Q_k(y)^n}{n!} e^{-Q_k(y)}, \tag{3.1}$$

where

$$Q_k(y) = I(y) \int_{E=E^k_{\min}}^{E^k_{\max}} \int_{E'} D(E, E') S(E') e^{-\mu(y,E')} dE' dE, \tag{3.2}$$

where $I(y)$ is the air-scan x-ray photon intensity at the yth source-detector pair, $S(E)$ is the normalized spectrum, $D(E, E')$ is the SRF, $\mu(y, E)$ is the integrated linear attenuation coefficient (LAC) at the yth source-detector pair and at energy E, and E^k_{min} and E^k_{max} are energy thresholds for energy bin k. Function (3.2) can be simplified by

$$D_k(E') = \int_{E=E^k_{\min}}^{E^k_{\max}} D(E, E') dE, \tag{3.3}$$

which is the BRF. The simplified (3.2) is

$$Q_k(y) = I(y) \int_E D_k(E) S(E) e^{-\mu(y,E)} dE. \tag{3.4}$$

Material decomposition model

Based on the BVM introduced in Chapter 2, each material can be decomposed into a linear combination of J basis materials with their respective linear attenuation spectrum $\mu_j(E)$, according to

$$\mu(y, E) = \sum_{j=1}^{J} \mu_j(E) l_j(y), \tag{3.5}$$

where $u_j(E)$ is the linear attenuation coefficient of basis material j at energy E, $l_j(y)$ is the line integral length of material j at the yth source-detector pair. The requirements for the selection of the basis materials are the same as stated in Chapter 2, and the same polystyrene and calcium chloride aqueous solution are chosen to be the basis materials.

Maximum likelihood estimation

Assuming the $d_k(y)$ form a set of independent Poisson random variables with means $Q_k(y)$, for each source-detector pair y, the basis-material line integral $l_j(y)$ can be estimated by minimizing the negative log-likelihood with arbitrary data $d_k(y)$, or

$$\min_{\{l_j(y), j=1,2,...,J\}} L\big(d_1(y), ..., d_K(y) | l_j(y)\big) = \min_{\{l_j(y), j=1,2,...,J\}} \sum_{k=1}^{K} \big[Q_k(y) - d_k(y) \log Q_k(y)\big], \tag{3.6}$$

where $Q_k(y)$ is defined in (3.2). By solving problem (3.6) independently for each y, the maximum likelihood estimates of the basis-material sinogram pixels $l_j(y)^{ML}$ are obtained. Because the number of basis materials is limited, Newton's method can be directly used to solve the minimization problem. After the basis-material sinograms are decomposed, conventional reconstruction using filtered backprojection is used to generated basis-material images.

3.5 Multi-energy joint statistical image reconstruction

In the previous section, the sinogram-based material decomposition method was introduced. However, this method generates large bias when the dose level is low as will be shown in the following simulation section. Because the FBP algorithm is a linear operator, this bias comes from the biased estimation of the basis-material sinograms. It has been shown that using the pre-log transmission data can achieve notably better quantitative accuracy than the post-log attenuation data in low-dose CT reconstruction [9]. Instead of reconstructing the basis-material images using FBP with biased basis-material sinograms, the joint statistical image reconstruction (JSIR) framework using the multi-energy alternating minimization algorithm is proposed to reconstruct the basis-material images directly from the transmission data.

3.5.1 Statistical model

In order to directly reconstruct the basis-material image c, equation (3.5) is rewritten as

$$\mu(y,E) = \sum_{j=1}^{J} \mu_j(E) \sum_x h(y|x) c_j(x), \tag{3.7}$$

where $h(y|x)$ is the CT system matrix element. Due to the presence of $h(y|x)$, each source-detector pair is not independent any more, and the new minimization problem becomes

$$\min_c L\big(d_1(y),...,d_K(y)|c\big) = \min_c \sum_y \sum_{k=1}^{K} \big[Q_k(y,c) - d_k(y)\log Q_k(y,c)\big], \tag{3.8}$$

where

$$Q_k(y,c) = I(y) \int_E D_k(E) S(E) e^{-\sum_{j=1}^{J} \mu_j(E) \sum_x h(y|x) c_j(x)} dE. \tag{3.9}$$

3.5.2 Alternating minimization

Instead of solving problem (3.8), the following problem is solved

$$\begin{aligned}\min_{p,q} F(p,q) &= \sum_y \sum_k \sum_E I(p_k(y,E)||q_k(y,E)) \\ &= \sum_y \sum_k \sum_E p_k(y,E)\log\frac{p_k(y,E)}{q_k(y,E)} - p_k(y,E) + q_k(y,E) \\ \text{s.t.} \quad &\sum_E p_k(y,E) = d_k(y) \\ &q_k(y,E) = I_k(y,E)e^{-\sum_j \mu_j(E)\sum_x h(y|x)c_j(x)},\end{aligned} \tag{3.10}$$

where $d_k(y)$ is the measurement of bin k for the yth source-detector-pair, and $I_k(y,E)$ is defined as the detector-response weighted spectrum, so

$$I_k(y,E) = I(y)D_k(E)S(E). \tag{3.11}$$

Theorem 3.1. *Solving problem* (3.10) *is equivalent to solving problem* (3.8).

The proof of this theorem is in Appendix A.3.

To minimize the cost function with respect to p, only the terms that contain p are considered

$$\min_p F_p(p) = \sum_{y,k,E} \Big[p_k(y,E)\log\frac{p_k(y,E)}{q_k(y,E)} - p_k(y,E) + q_k(y,E)\Big], \tag{3.12}$$

$$\text{s.t.} \sum_E p_k(y,E) = d_k(y).$$

After solving for the Lagrangian multiplier, the minimizer of problem (3.12) is

$$p_k(y,E) = \frac{q_k(y,E)}{\sum_E q_k(y,E)} d_k(y). \tag{3.13}$$

From (3.10), q is a function of c. Minimizing the cost function with respect to q is equivalent to minimizing the cost function with respect to c. Only the terms that contain c are considered to minimize the cost function with fixed p, or

$$\begin{aligned}\min_c F_c(c) &= \sum_{y,k,E} -p_k(y,E)\log q_k(y,E) + q_k(y,E) \\ &= \sum_{y,k,E} \Big[I_k(y,E) e^{-\sum_j \mu_j(E) \sum_x h(y|x) c_j(x)} + p_k(y,E) \sum_j \mu_j(E) \sum_x h(y|x) c_j(x) \Big].\end{aligned} \tag{3.14}$$

In (3.14), the values of $c_j(x)$ are coupled with each other, so the convex decomposition Lemma 2.1 is used to decouple them. The resulting surrogate function is

$$\hat{F}_c(c) = \sum_{y,k,E} \left[\hat{q}_k(y,E) \sum_{j,x} \frac{\mu_j(E) h(y|x)}{Z_j(x)} e^{-Z_j(x)\left(c_j(x) - \hat{c}_j(x)\right)} + p_k(y,E) \sum_j \mu_j(E) \sum_x h(y|x) c_j(x) \right], \tag{3.15}$$

where

$$\hat{q}_k(y,E) = I_k(y,E) e^{-\sum_{j,x} \mu_j(x) h(y|x) \hat{c}_j(x)}, \tag{3.16}$$

$$Z_j(x) = \sum_{y,E} \mu_j(E) h(y|x). \tag{3.17}$$

$\hat{c}_j(x)$ is the current estimate of $c_j(x)$ and Z is a parameter that guarantees convergence. For the surrogate function (3.15), the derivative with respect to $c_j(x)$ is

$$\frac{\partial \hat{F}_c}{c_j(x)} = \sum_{y,k,E} \Big[-\hat{q}_k(y,E) \mu_j(E) h(y|x) e^{-Z_j(x)\left(c_j(x) - \hat{c}_j(x)\right)} + p_k(y,E) \mu_j(E) h(y|x) \Big]. \tag{3.18}$$

Setting this derivative equal to zero gives

$$e^{-Z_j(x)\left(c_j(x)-\hat{c}_j(x)\right)} = \frac{\sum_{y,k,E} p_k(y,E)\mu_j(E)h(y|x)}{\sum_{y,k,E} q_k(y,E)\mu_j(E)h(y|x)}. \tag{3.19}$$

The final $c_j(x)$ update equation is

$$c_j(x) = \hat{c}_j(x) - \frac{1}{Z_j(x)} \log\left(\frac{\sum_{y,k,E} p_k(y,E)\mu_j(E)h(y|x)}{\sum_{y,k,E} q_k(y,E)\mu_j(E)h(y|x)}\right). \tag{3.20}$$

The equation above gives a closed-form update of the basis-material images. However, with low-dose levels, the image quality will be degraded by high noise. Regularization is used to suppress noise. Again, a Huber-type penalty is selected with the form [11, 12]

$$R(c) = \sum_j \sum_x \sum_{\tilde{x}\in\mathcal{N}_x} w_{x\tilde{x}}\phi\big(c_j(x) - c_j(\tilde{x})\big), \tag{3.21}$$

where

$$\phi(c) = \frac{1}{\delta^2}\big(\delta|c| - \log(1+\delta|c|)\big). \tag{3.22}$$

$\mathcal{N}_x$ is the set of neighborhood pixels of pixel x. The weights $w_{x\tilde{x}}$ are equal to the inverse distances between $\tilde{x}$ and x. An advantage of this penalty choice is that it is twice continuously differentiable at 0. δ is a tuning parameter which defines the range where $\phi(c)$ performs like a quadratic function $|c|^2$, and outside this region, $\phi(c)$ performs like $|c|$. In this chapter, δ is set to be 20.

The regularized MEAM cost function is

$$\min_{c} \hat{F}_c(c) = \sum_{y,k,E} \Bigg[\hat{q}_k(y,E) \sum_{j,x} \frac{\mu_j(E) h(y|x)}{Z_j(x)} e^{-Z_j(x)\left(c_j(x) - \hat{c}_j(x)\right)} + p_k(y,E) \sum_j \mu_j(E) \sum_x h(y|x) c_j(x) \Bigg] + \lambda R(c). \tag{3.23}$$

Due to the presence of the penalty term, there is no longer a closed-form update. Instead, Newton's method is utilized to update the basis-material images. The penalty term $\phi\big(c_j(x) - c_j(\tilde{x})\big)$ couples the values of $c_j(x)$ and $c_j(\tilde{x})$, so the convex decomposition lemma is used again to decouple them. The forms of the first order and second order derivatives of the decoupled cost function are

$$\frac{\partial \hat{F}_c}{\partial c_j(x)} = \sum_{y,k,E} \Big[- \hat{q}_k(y,E) \mu_j(E) h(y|x) e^{-Z_j(x)\left(c_j(x) - \hat{c}_j(x)\right)} + p_k(y,E) \mu_j(E) h(y|x) \Big] + \lambda \sum_{\tilde{x} \in \mathcal{N}_x} w_{x\tilde{x}} \frac{2c_j(x) - \hat{c}_j(x) - \hat{c}_j(\tilde{x})}{1 + \delta |2c_j(x) - \hat{c}_j(x) - \hat{c}_j(\tilde{x})|} \tag{3.24}$$

and

$$\frac{\partial \hat{F}_c^2}{\partial^2 c_j(x)} = \sum_{y,k,E} \hat{q}_k(y,E) \mu_j(E) h(y|x) Z_j(x) e^{-Z_j(x)\left(c_j(x) - \hat{c}_j(x)\right)} + \lambda \sum_{\tilde{x} \in \mathcal{N}_x} w_{x\tilde{x}} \frac{2}{\big(1 + \delta |2c_j(x) - \hat{c}_j(x) - \hat{c}_j(\tilde{x})|\big)^2}, \tag{3.25}$$

The pseudo-code for the MEAM algorithm for PCCT is presented in Algorithm 2.

Algorithm 2: Multi-energy alternating minimization algorithm

Initialize $c_j(x)$, and compute the corresponding $\hat{q}_k(y, E)$ and $\hat{p}_k(y, E)$.

if without penalty **then**

  for $n = 0$ to $N - 1$ **do**

    foreach j **do**

      Set $c_j(x)^{(n)}$ as the current estimate.

      Update $\hat{q}_k(y, E)$ with (3.16).

      Update $\hat{p}_k(y, E)$ with (3.13).

      Update $c_j(x)^{(n+1)}$ with (3.20).

    end

  end

end

if with penalty **then**

  for $n = 0$ to $N - 1$ **do**

    foreach j **do**

      Set $c_j(x)^{(n)}$ as the current estimate.

      Update $c_j(x)^{(n+1)}$ with Newton's method, and the first order and second order derivatives are from (3.24) and (3.25).

    end

  end

end

3.6 Patient phantom simulation

3.6.1 Simulation setup

A virtual patient phantom was used to compare the quantitative accuracy of the sinogram-based decomposition method and the MEAM algorithm. The patient phantom was a pelvis abdomenal slice from the ICRP adult female phantom [24]. The proprietary geometry of the Philips Brilliance Big Bore CT scanner (Philips Medical Systems, Cleveland, OH) with 816 detectors per detector row and 1320 source positions per rotation was used in this simulation. The photon-counting detector SRF and BRF were generated from the Photon Counting Toolkit (PcTK)[43]. The radius of the electronic charge cloud is $\rho_0 = 24\,\mu\text{m}$, the electronic noise parameter $\delta_e = 4\,\text{keV}$, the PCD size in channel-row $d_{pix} = 225\,\mu\text{m}$, the PCD thickness $d_z = 1600\,\mu\text{m}$, and 5 measurement bins with thresholds $E_1 = 20\,\text{keV},\ E_2 = 40\,\text{keV},\ E_3 = 55\,\text{keV}, E_4 = 70\,\text{keV}, E_5 = 90\,\text{keV}$ were generated. Charge sharing and electronic noise were simulated, and the pulse pileup effect was not taken into consideration. The BRFs of 5 energy bins are presented in Figure 3.1. In this simulation, a source spectrum of 120 kVp was used, and the dose levels were set to be 30 mAs, 50 mAs, 100 mAs, and 200 mAs to test the quantitative performance with different dose levels.

3.6.2 Performance assessment

The basis-material images $c_j(x)$ were reconstructed with both the sinogram-based decomposition method and the MEAM algorithm. In this simulation, polystyrene and 23% $CaCl_2$ aqueous solution were used as the basis materials, which was proven to parameterize the LAC to within 1-2% accuracy [5, 18, 19]. The LAC were computed based on the BVM

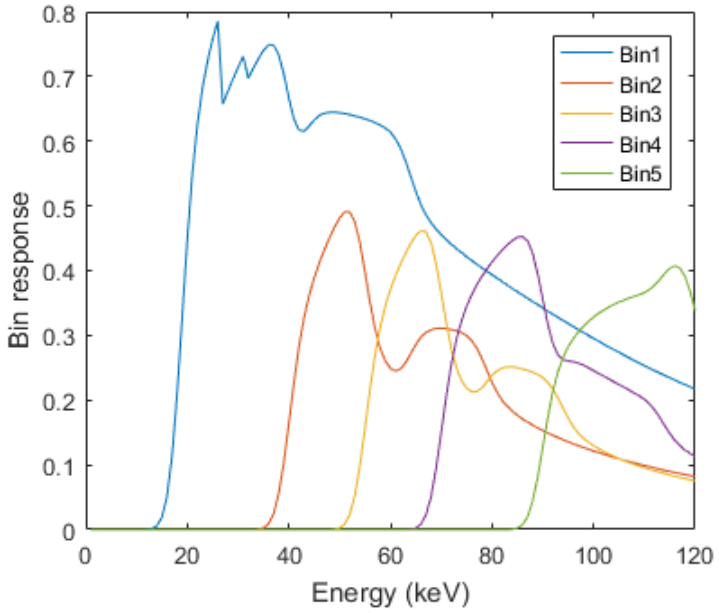


Figure 3.1: The bin response functions of 5 bins from PcTK

model according to

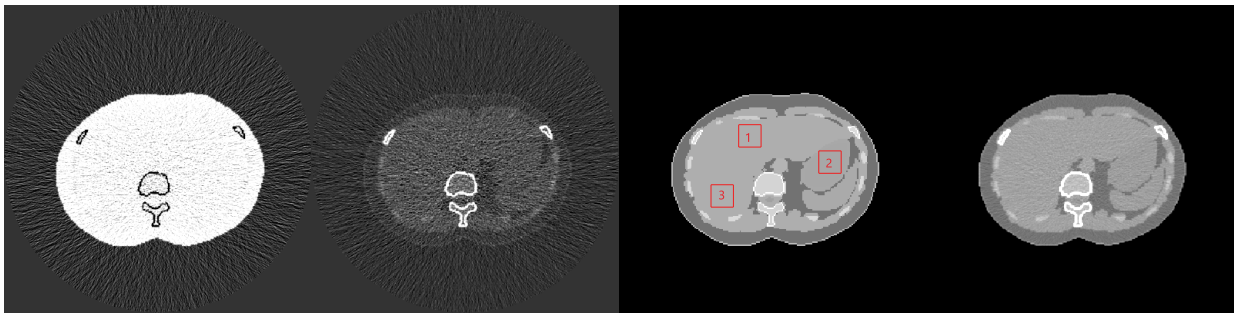
$$\mu(x, E) = \mu_1(E)c_1(x) + \mu_2(E)c_2(x), \tag{3.26}$$

where c_1 and c_2 are the basis-material images of polystyrene and $CaCl_2$ aqueous solution, and μ_1 and μ_2 are the corresponding LAC. The reconstructed LAC were compared with the reference ground-truth values. The ground-truth values of water and other sample materials were from the NIST XCOM database [20]. To better show the results quantitatively, the CT numbers were computed based on

$$\mathrm{HU}(E) = 1000\left(\frac{\mu(E)}{\mu_{\mathrm{water}}(E)} - 1\right). \tag{3.27}$$

3.6.3 Sinogram-based decomposition results

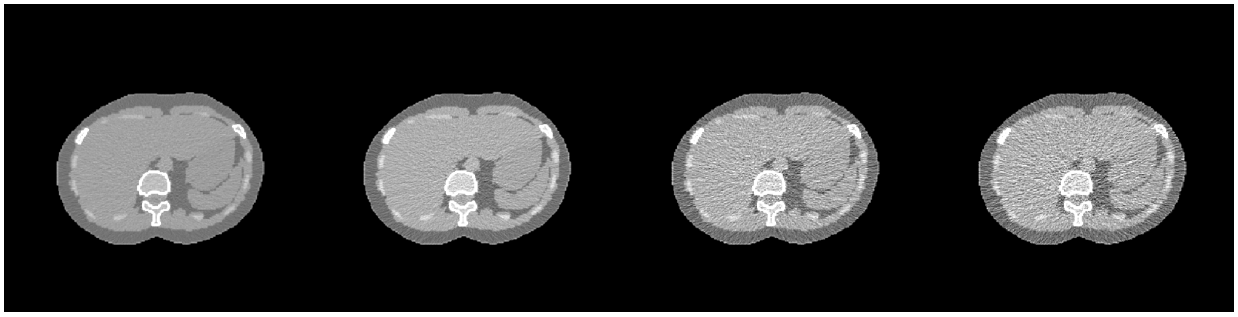
In this subsection, the results using the sinogram-based decomposition method are shown. The result with 200 mAs dose level is presented in Figure 3.2. In Figure 3.2(a) and (b), the decomposed basis images of polystyrene and calcium chloride aqueous solution are presented. The ground truth image of the LAC at 60 keV is shown in Figure 3.2(c), and the reconstructed LAC image is shown in Figure 3.2(d). At 200 mAs, the reconstructed LAC



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

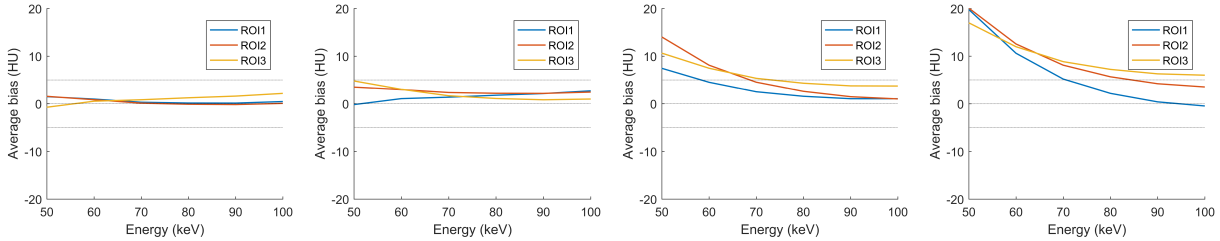
Figure 3.2: With 200 mAs dose level, (a) the decomposed image of polystyrene, (b) the decomposed image of calcium chloride aqueous solution. The viewing windows are [-0.2 1]. (c) The ground truth image of the ICRP phantom linear attenuation coefficient at 60 keV, (d) the reconstructed image of the linear attenuation coefficient with the sinogram-based method at 60 keV. The viewing windows are [-300 300] HU.

image shows low artifacts and noise. To further study the performance of the sinogram-based decomposition method, the LAC image were reconstructed with 30 mAs, 50 mAs, and 100 mAs. These images, combined with the 200 mAs reconstruction, are shown in Figure 3.3. A direct observation is that with decreased dose level, the noise level increases and streak artifacts begin to emerge. To quantitatively assess the performance of the sinogram-based decomposition method, three homogeneous regions of interest (ROI) were selected to compute the average bias spectra inside these regions, which are highlighted by the red boxes in Figure 3.2(c). In Figure 3.4, the biases spectra of these three homogeneous regions are



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

Figure 3.3: The reconstructed LAC images with various dose levels: (a) 200 mAs, (b) 100 mAs, (c) 50 mAs, and (d) 30 mAs. The viewing windows are [-300 300] HU.



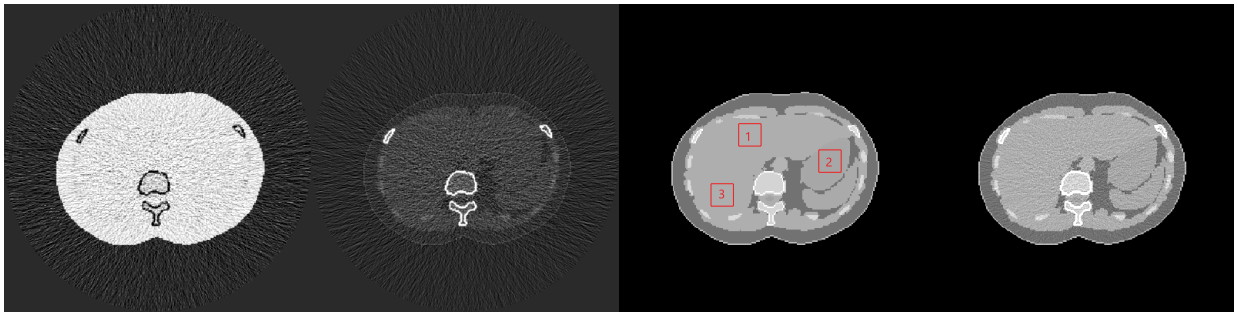
(a) (b) (c) (d)

Figure 3.4: The average LAC biases of ROIs with various dose levels: (a) 200 mAs, (b) 100 mAs, (c) 50 mAs, (d) 30 mAs.

presented. When the dose level is as high as 200 mAs, as shown in Figure 3.4(a), the bias is within 2 HU across the spectrum, in which case the bias can be ignored. In the 100 mAs case, the biases of three ROIs have a minor positive shift, but still stay within a 5 HU limit. If the dose level is reduced to 50 mAs, significant positive bias in all three ROIs can be seen and they exceed 5 HU when the energy is below 60 keV. In the ultra low-dose case of 30 mAs, the positive bias further increases. In this simulation, it has been demonstrated that with the sinogram-based decomposition method, when the dose level is reduced, a positive bias emerges.

3.6.4 MEAM results

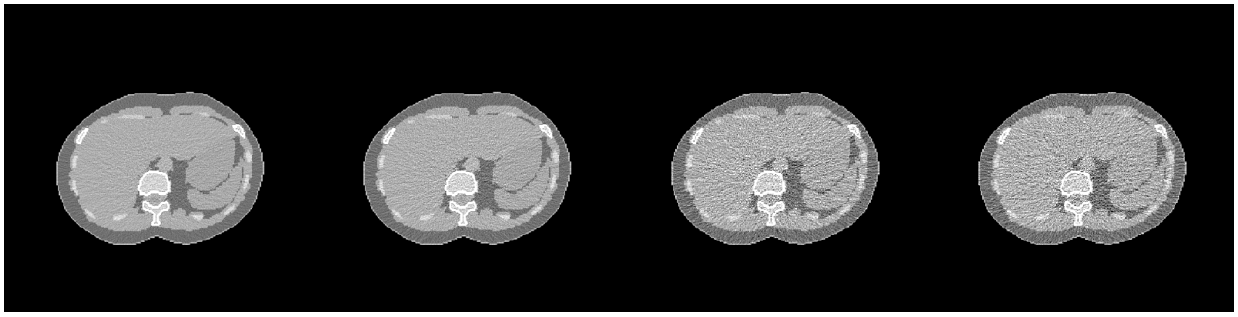
The unregularized MEAM algorithm results using the same simulation data are shown in Figure 3.5, Figure 3.6, and Figure 3.7. Figure 3.5(a) and (b) show the decomposed basis-material images. Compared with the sinogram-based decomposition method results in Figure 3.2(a) and (b), MEAM suppresses the streak artifacts in the basis-material images. The



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

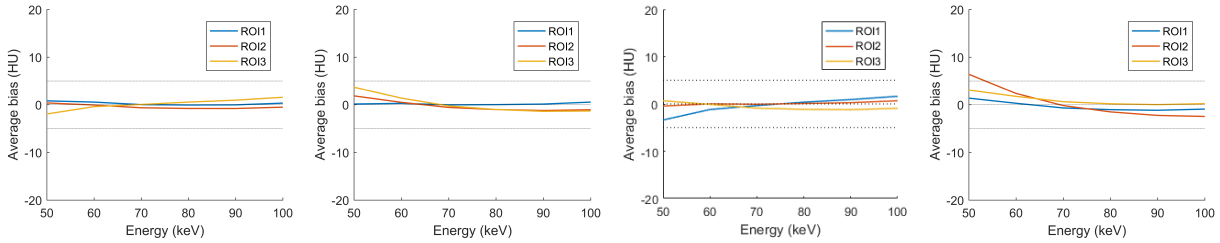
Figure 3.5: With a 200 mAs dose level, (a) the decomposed image of polystyrene, (b) the decomposed image of calcium chloride aqueous solution. The viewing windows are [-0.2 1]. (c) The ground truth image of the ICRP phantom linear attenuation coefficient at 60 keV, (d) the reconstructed image of the linear attenuation coefficient with MEAM without regularization at 60 keV. The viewing windows are [-300 300] HU.

reconstructed LAC image at 60 keV is presented in Figure 3.5(d), which shows low noise and artifacts. The reconstructed LAC image at 60 keV with different dose levels are shown in Figure 3.6, and the corresponding three homogeneous regions bias spectra are shown in Figure 3.7. With high-dose levels (200 mAs and 100 mAs), the MEAM generated nearly unbiased estimates just like the sinogram-based decomposition method. However, when the dose level decreased to 50 mAs and 30 mAs, the sinogram-based method generated large biases across the spectrum, and the MEAM results show that the bias can still be constrained to within a very tight window of ± 5 HU.



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

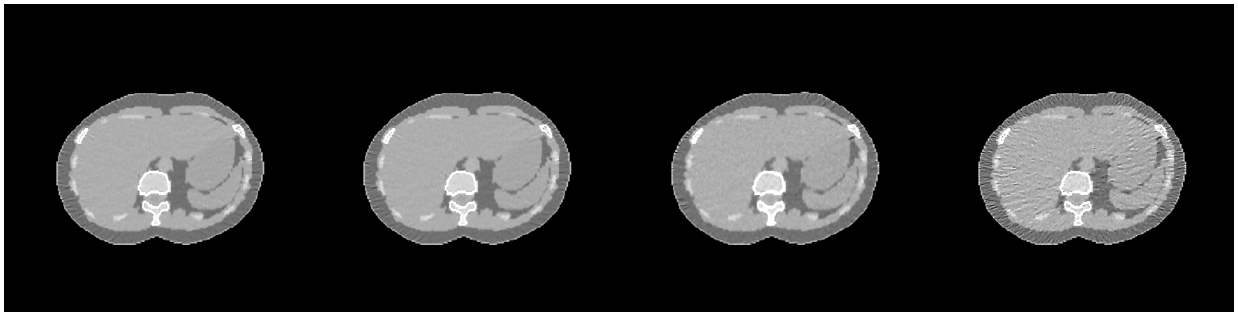
Figure 3.6: The reconstructed LAC images with various dose levels: (a) 200 mAs, (b) 100 mAs, (c) 50 mAs, (d) 30 mAs. The viewing windows are [-300 300] HU. All images were reconstructed without regularization.



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

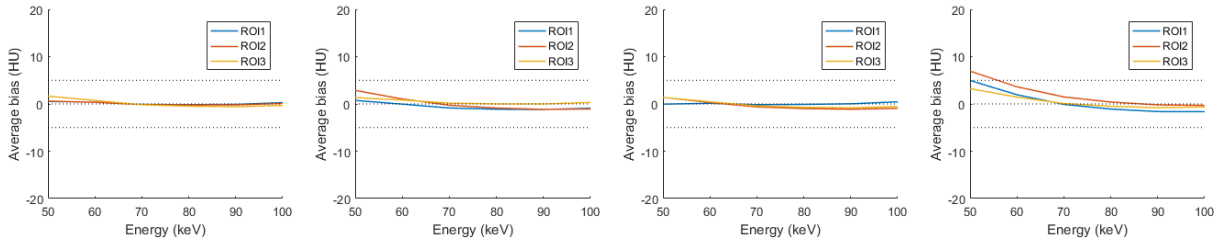
Figure 3.7: The average LAC biases of ROIs with various dose levels: (a) 200 mAs, (b) 100 mAs, (c) 50 mAs, and (d) 30 mAs.

It has been shown that the unregularized MEAM algorithm can dramatically reduce the bias even in an ultra-low-dose scenario; however, the high noise with low dose is still a major concern. To suppress the noise, the image-domain neighborhood penalty was used. The reconstructed LAC images at 60 keV are shown in Figure 3.8, and their corresponding homogeneous regions bias spectra are shown in Figure 3.9. The noise clearly gets reduced without sacrificing edge resolution, and the bias spectra are still kept close to 0.



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

Figure 3.8: The reconstructed LAC images with various dose levels: (a) 200 mAs, (b) 100 mAs, (c) 50 mAs, (d) 30 mAs. The viewing windows are [-300 300] HU. All images were reconstructed with the same intensity-normalized regularization.



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

Figure 3.9: The average LAC biases of ROIs with various dose levels: (a) 200 mAs, (b) 100 mAs, (c) 50 mAs, (d) 30 mAs.

3.7 Discussion

From the simulation results, it can be seen that the sinogram-based method generates large bias in the reconstructed LAC images when the dose level is low. Because FBP is a linear operator, then the bias corresponds to the biased decomposed basis-material sinograms. For a single source-detector pair, assuming the scanned object is a uniform calcium chloride aqueous solution with fixed length l and the x-ray beam is mono-energetic with energy E, the ground truth of the line integral of LAC is $\mu(E) = \mu_{CaCl_2}(E)l$. From the discussion in the previous chapter, the estimated value of the reconstructed LAC has a positive bias, which

is $\mathrm{E}(\hat{\mu}(E)) > \mu(E)$. From the discussion in the dual-energy CT scan chapter and the proof in Appendix A.1, the sinogram-based decomposition methods generate biased estimates and the bias decreases with decreased dose level.

A Monte-Carlo simulation was used to show the relationship between the bias and dose level for sinogram-based decomposition methods. The same spectrum and BRF as in the above simulation were used. The basis materials were still chosen to be polystyrene and calcium chloride aqueous solution. The scanned object was 250 mm of polystyrene. With 1 million Monte-Carlo trials for each dose level from 20 mAs to 200 mAs, the relative estimation error of the LAC is shown in Figure 3.10. It is clear that with the increase of the dose level, the relative estimation bias monotonically decreases.

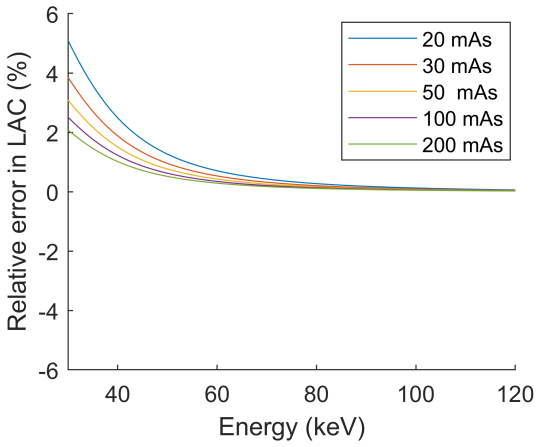


Figure 3.10: The Monte-Carlo simulation of the relationship between relative error and dose level.

Aside from photon starvation due to low-dose levels, another factor that leads to the bias, as stated in Chapter 2, is the lack of accounting for the system geometry in the sinogram-decomposition method. The basis-material sinograms are decomposed by solving problem

(3.6) independently for each source-detector pair y. The assumption here is that each basis-material sinogram pixel is independent of each other. However, this assumption is not valid due to the presence of the forward- projection model

$$l_j(y) = \sum_x h(y|x)c_j(x), \tag{3.28}$$

such that some source-detector pairs may share a common pixel x. In the MEAM algorithm, the forward model is directly incorporated, and basis-material images are reconstructed from the statistical model.

3.8 Summary

Like the dual-energy scan introduced in the previous chapter, the PCCT system is also able to derive energy-related information by material decomposition. The conventional PCCT sinogram-based decomposition also suffers a large bias due to photon starvation and lack of accounting for the system geometry. The MEAM algorithm modifies the statistical model by incorporating the system geometry and can effectively eliminate the systematic bias in the sinogram-based decomposition method even in the ultra-low-dose scenario.

Chapter 4

Auxiliary sinogram alternating minimization (ASAM) algorithm

4.1 Highlights

In this chapter, the possibility of enforcing a penalty constraint on the post-log attenuation which traditionally serves as the data input will be explored. In this way, the classic interpolation method to reduce metal artifacts can be incorporated in the iterative algorithm framework.

4.2 Motivation

Modern x-ray CT imaging is able to produce high resolution images and is applied in many areas like medical imaging [4, 45, 46] and security imaging [47] [48]. The attenuation coefficients of materials are reconstructed to get an anatomical image of the patient for medical applications or the contents of luggage for security applications. However, when there exist

dense or metal materials in the scanned object, the image quality may suffer severe artifacts. Several categories of methods to reduce metal artifacts have been developed [49]. Among them, the sinogram-based preprocessing method [50] and iterative reconstruction algorithms [51] are two effective ones. However, there is still no systematic way to take advantage of both methods. To deal with this problem, we developed a constrained optimization scheme to synchronize these two methods and further improve metal-artifact reduction. We use penalized discrepancies in the measured sinogram, the auxiliary sinogram, and image domain to accomplish this, adapting the dual-domain optimization method from Chen *et al.* [23] to address metal-artifact reduction.

4.3 Notation

In this section, the notations used in this chapter are summarized

- d: measured transmission data.
- μ: reconstructed image.
- S: auxiliary sinogram.
- I: air-scan source intensity.
- H: system matrix.
- C: set of sinogram pixels not affected by the metal object.
- $\tilde{C}$: set of sinogram pixels affected by the metal object.
- i: source-detector pair index.

- j: image pixel (voxel) index.

4.4 Background

4.4.1 Causes of metal artifacts

Streak artifacts due to metal are prevalent in CT images. They are caused by several mechanisms. Some are due to the metal itself; some are related to limited scanner resolution and the motion of the scanned object.

Beam hardening

Beam hardening is a common phenomenon in x-ray CT scans. Instead of the ideal monoenergetic x-ray beam, the realistic beam emitted from the source tube has a spectrum spread over a range of energies. When the x-ray beam penetrates an object, the lower energy part is more likely to be absorbed compared to the higher energy photons. Thus, the effective energy of the x-ray beam increases, i.e., the beam is hardened. When the image is reconstructed without beam hardening correction, this mismatch results in the so-called *interpetrou lucency artifact* [52].

This problem is more severe for materials with relatively high atomic number such as bones or metal endoprostheses in patients. Compared with low-atomic-number materials (i.e., soft tissues), materials with high atomic number have a steeper decrease in attenuation coefficient with increased energy. Even if the scanned object is homogeneous, the different path length of x-ray beams will cause the cupping effect as shown in Figure 4.2. The other resulting

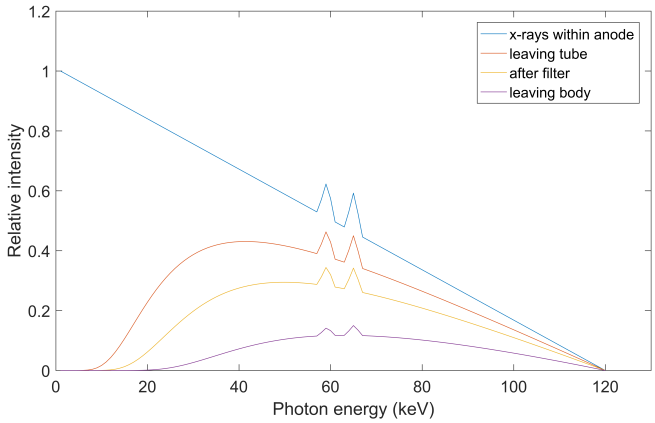
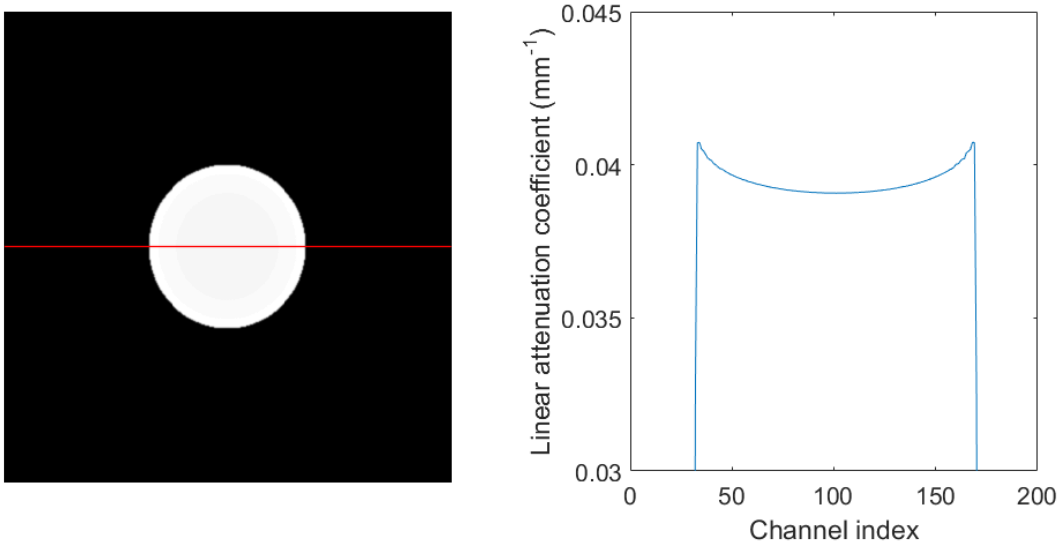


Figure 4.1: The x-ray spectra in different stages in CT scans. The spectrum detected by the detectors has the highest effective energy.

artifacts include dark streaks connecting high attenuation objects, dark shades around metal material, and streaks coming out of the high attenuation objects [53].



(a) (b)

Figure 4.2: Cupping effect due to beam hardening.

Scatter

In x-ray CT, we assume each x-ray photon travels a straight path through the scanned object and hits the detector. However, when an x-ray photon is scattered, it deviates from its original path and contributes to the measurement at other detectors. When the primary photon count at a detector is low, the scatter-to-primary ratio (SPR) increases, and scatter becomes the main source of artifacts. For highly attenuated x-ray beams, scatter causes more photons to be detected than expected. Therefore, in the attenuation sinogram, these metal traces have much lower values than the ground truth [54]. If the FBP algorithm, which is a linear operator, is used to reconstruct the image, the underestimation in the attenuation sinogram will result in an underestimation in the reconstructed image. The corresponding artifacts look similar to those caused by beam hardening with dark streaks between dense materials and dark shades around them.

Poisson noise

The measurement noise at the detector can be modeled by a Poisson process. When the primary photon count is low, this Poisson noise may cause notable artifacts. The artifacts may contain dark and bright streaks in the direction of greatest attenuation and result in low contrast in the soft tissues. One possible way to suppress the Poisson noise is to increase the x-ray source exposure. In modern CT scanners, when dense or metal objects are in the field of view of projection, a temporary x-ray tube current increase is applied by adaptive x-ray tube modulation [53].

Other factors

Other factors that may result in artifacts include miscalibration between x-ray sources and detectors, the limited resolution of the CT scanner that therefore cannot have a clear boundary identification for high contrast objects, and motion during a CT scan.

4.4.2 Current methods to reduce metal artifacts

Various types of metal-artifact reduction (MAR) have been developed [55, 56, 57, 50, 58, 59]. They can generally be classified into two categories: the preprocessing method and the iterative algorithm.

Preprocessing method

The preprocessing method has been widely used for more than 30 years. The basic idea is that the measurement with the greatest attenuation is heavily influenced by the factors we have stated above, so instead of using this part of the measurement, surrogate data is manually generated and used to help with the image reconstruction. The simplest example is linear interpolation [55]. First, the dense or metal part is identified from an initial reconstructed image with FBP [60], then we forward project this part to get its corresponding partial sinogram, and the nonzero entries in this partial sinogram are noted as metal traces to be interpolated. Then in the original sinogram, we linearly interpolate the metal traces and get a corrected sinogram. Thus we can reconstruct a new image with reduced metal artifacts. In this way, we can reduce the dark and bright streaks and shadows around metal objects. However, there is a clear drawback of this method, because we completely ignore

the metal traces and this loss of information will result in a blurring of edges in the image. The harsh separation between metal and non-metal traces may introduce new streaks in the final reconstructed image. Several improved methods have been proposed based upon this basic linear interpolation method in order to get more accurate surrogate data, for example, smooth interpolation [56], the segmentation and reprojection method [57, 59], nonlinear sinogram smoothing [58], and normalized metal-artifact reduction (NMAR) [50]. However, in all of these preprocessing methods, the problem of loss of information remains. Another concern is that these methods do not compensate for the underestimation caused by high SPR in the metal-affected channels. In this chapter, we use NMAR as the default preprocessing method for performance comparison.

Iterative algorithm

Instead of using the attenuation sinogram, we use the full photon count measurement and model it as Poisson random variables [61, 62]. Compared with the preprocessing methods, the usage of the pre-log measurement can achieve notably better quantitative accuracy and reduce the artifacts caused by low-dose measurements [9]. Also, the iterative algorithm is a completely different method for reducing the metal artifacts compared than the preprocessing methods. One clear advantage is that it uses all of the measured data. The iterative algorithm addressed in this chapter is the *alternating minimization* (AM) algorithm[4][51]. Another advantage of the iterative algorithm is that we can include prior knowledge like smoothness in the region of interest to turn the AM algorithm into a penalized optimization problem to suppress the noise in the image [63, 64]. However, the idea of interpolating the metal traces was not applied in the AM algorithm.

4.5 Auxiliary sinogram alternating minimization algorithm

Both the preprocessing method and the AM algorithm can reduce metal artifacts and have their limitations and disadvantages. A natural idea is to combine these two methods systematically. In the present study, we introduce the auxiliary sinogram domain such that the interpolation method can be applied in the iterative reconstruction algorithm without changing the raw measured data. A weighting parameter on the metal traces is introduced to emphasize the data that are not affected by the metal and down-weight the data affected by the metal. In this way, we avoid loss of information, in which the data affected by the metal is fully ignored [51], and the underestimation caused by high SPR is compensated for by the use of the auxiliary sinogram.

4.5.1 Data model

Let $d_{i,i\in[1...M]}$ denote the transmission measurement at the ith source-detector pair. Following Lange and Carson [44], we model the data as mutually independent Poisson distributions

$$Pr[d_i = n] = \frac{Q_i^n}{n!} e^{-Q_i} \tag{4.1}$$

for $n = 0, 1, 2, ...$ and $i \in 1, 2, ..., M$, in which

$$Q_i = I_i \exp\left(- \sum_j h_{ij}\mu_j \right), \tag{4.2}$$

where I_i is the source flux of the ith source-detector pair, j is the image pixel index, and h_{ij} is the scanner's point spread function (PSF) generated from the system geometry.

In AM, we reconstruct the image directly from the measurement,

$$\mathbf{D} \xrightarrow{\text{AM}} \mathbf{U},$$

while the preprocessing methods are sinogram-based and use linear methods like FBP to reconstruct the image

$$\mathbf{L} \xrightarrow{\text{FBP}} \mathbf{U},$$

where $L_i = -\log(\frac{d_i}{I_i})$ represents the attenuation sinogram for the ith source-detector pair. The difference in the input domain is one of the obstacles we mentioned earlier that prevents the use of preprocessing method and the AM algorithm at the same time.

4.5.2 Statistical model

A new parameter **S** called the *Auxiliary Sinogram* is introduced, which serves as an intermediate state between the raw transmission sinogram d and the reconstructed image μ.

$$\mathbf{D} \longrightarrow \mathbf{S} \longleftrightarrow \mathbf{U}.$$

The term of the forward projection of the image, $\sum_j h_{ij}\mu_j$, is replaced by a single auxiliary sinogram pixel S_i. From equation (4.1) and (4.2), the connections between d_i, S_i, and μ are shown as

$$Pr[d_i = n] = \frac{Q(S_i)^n}{n!} e^{-Q(S_i)}, \tag{4.3}$$

where

$$Q(S_i) = I_i \exp(-S_i), \tag{4.4}$$

$$S_i = \sum_j h_{ij}\mu_j. \tag{4.5}$$

I-divergence is used to measure the discrepancy between the measurement data and auxiliary sinogram and form the reconstruction problem as

$$\min_S \sum_i I\big(d_i||Q(S_i)\big) = \min_S \sum_i \left[d_i \log \frac{d_i}{Q(S_i)} - d_i + Q(S_i)\right] \tag{4.6}$$

$$\text{s.t. } S_i = \sum_j h_{ij}\mu_j.$$

To solve this problem, instead of forcing the equality constraint, the constraint $S_i = \sum_j h_{ij}\mu_j$ is softened and turned into a penalty term. As described in the previous study of the line integral alternating minimization (LIAM) algorithm by Chen *et al.* [23], instead of reconstructing the image by minimizing the discrepancy between the measurement data and the mean estimates, a tunable discrepancy between the auxiliary sinogram and the measured sinogram is added. Using I-divergence as a discrepancy measure, we reformulate our problem as

$$\min_{S,\mu} F(S,\mu) = \min_{S,\mu} \sum_i I\big(d_i||Q(S_i)\big) + \beta \sum_i I\bigg(S_i|| \sum_j h_{ij}\mu_j\bigg), \tag{4.7}$$

where the weight β controls the trade-off between accuracy for the measurements and the auxiliary sinogram. Theoretically, if we let $\beta \to \infty$, problem (4.7) is equivalent to problem (4.6). So by this dual-domain minimization problem we are able to overcome the obstacle of different input domains for preprocessing methods and the AM algorithm. Also, it is easy to apply the idea of emphasizing the accurate data traces that are not affected by metal and down-weight the metal traces. As in the preprocessing method, the metal pixels are

segmented from an initial image reconstructed by FBP. Then these metal pixels are forward projected to generate a metal-only sinogram. The non-zero sinogram pixels in the metal only sinogram are defined as $\tilde{C}$. The non-metal sinogram pixels are denoted as C. The generation of set C and $\tilde{C}$ are illustrated in Figure 4.3.

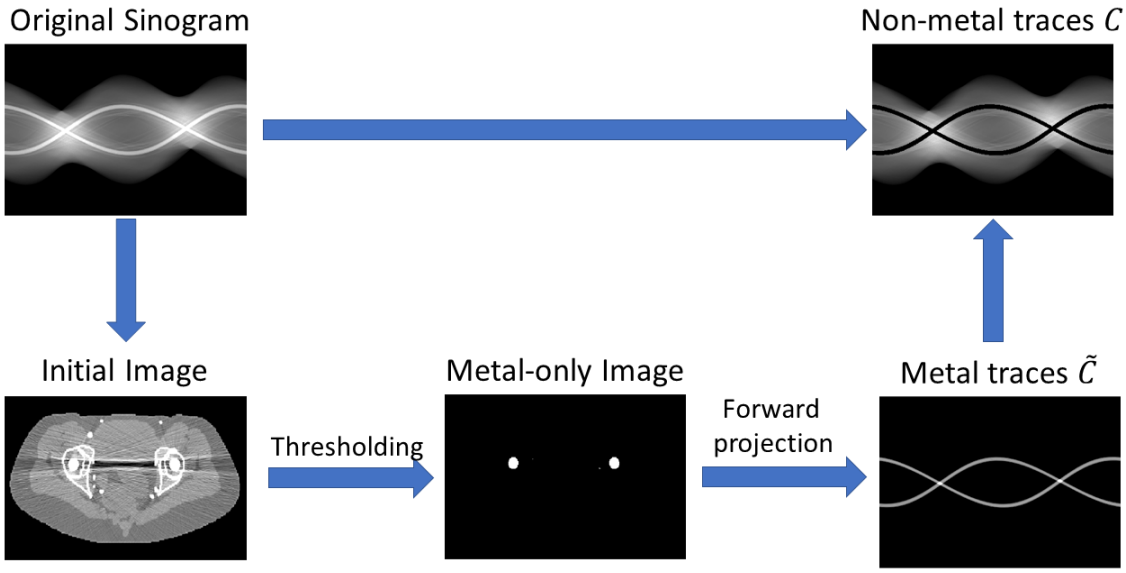


Figure 4.3: From the raw data, an uncorrected image is generated. By thresholding, the metal pixels are segmented. After the forward projection, the set of metal traces $\tilde{C}$ is defined. The other sinogram pixels are defined as non-metal traces C.

By this separation, problem (4.7) is reformulated as

$$\min_{S,\mu} F(S,\mu) = \min_{S,\mu} \sum_i \alpha_i I\big(d_i || Q(S_i)\big) + \beta \sum_i I\left(S_i || \sum_j h_{ij}\mu_j\right), \tag{4.8}$$

where

$$\alpha_i = \begin{cases} 1 & i \in C \\ \alpha_0 & i \in \tilde{C}, \end{cases}$$

where α_0 is a weighting parameter that controls the balance between metal sinogram pixels and non-metal sinogram pixels. For the inaccurate metal sinogram pixels, a low weight is imposed; and if we make $\alpha_0 \to 0$, the metal traces are completely ignored just as in [51].

4.5.3 Regularized ASAM

S domain regularization

In the previous section, the idea of auxiliary sinogram was introduced to achieve the goal of giving different weights for metal traces and non-metal traces, and also using the transmission measurement as the input. However, the preprocessing method is not involved in ASAM directly, and we implement this preprocessing in ASAM by putting a constraint on the S domain, so now our problem becomes a constrained minimization problem,

$$\min_{S,u} F(S,u) = \min_{S,u} \sum_i \alpha_i I\big(d_i || Q(S_i)\big) + \beta \sum_i I\bigg(S_i || \sum_j h_{ij} u_j\bigg) \tag{4.9}$$
$$\text{s.t. } R(S) \leq \xi,$$

where $R(S)$ accounts for different forms of constraints corresponding to methods like linear interpolation, polynomial interpolation and smooth interpolation. We choose $R(S)$ to be of the form

$$R(S) = \sum_{i \in \tilde{C}} \sum_{k \in \mathcal{N}_i} w_{ik} \phi(S_i - S_k : \delta), \tag{4.10}$$

where

$$\phi(x:\delta) = \frac{1}{\delta^2}\big(\delta|x| - \log(1+\delta|x|)\big). \tag{4.11}$$

This Huber-type penalty has been introduced in Chapters 2 and 3. In this chapter, the δ of the auxiliary sinogram domain regularization $R(S)$ is set to be 50. $\mathcal{N}_i$ is the set of neighborhood pixels around pixel i in the auxiliary sinogram. The weights w_{ik} control the relative contribution of each neighboring pixel of i. Here, we choose an adjacent pixel neighborhood around i and weights w_{ik} are equal to the inverse distance between pixels i and k. This constraint only applies to metal traces, that is only to $S_i, i \in \tilde{C}$ instead of the whole auxiliary sinogram S.

Just like what we did for problem (4.6), we turn this constrained minimization problem into a regularized minimization problem

$$\begin{aligned}\min_{S,\mu} F(S,\mu) = &\min_{S,\mu} \sum_i \alpha_i I\big(d_i || Q(S_i)\big) + \beta \sum_i I\left(S_i || \sum_j h_{ij}\mu_j\right) \\ &+ \gamma \sum_{i\in\tilde{C}} \sum_{k\in\mathcal{N}_i} w_{ik}\phi(S_i - S_k : \delta).\end{aligned} \tag{4.12}$$

When we increase the parameter $\gamma \to \infty$, the solution for problem (4.12) is equivalent to problem (4.9) with $\xi = 0$.

μ domain regularization

Because we have two domains to optimize, the μ domain can also be regularized. In fact, this is more intuitive because a smooth image is desirable. Using a similar weighted total

variation penalty, we get the following minimization problem

$$\begin{aligned}\min_{S,\mu} F(S,\mu) = &\min_{S,\mu} \sum_i \alpha_i I\big(d_i||Q(S_i)\big) + \beta \sum_i I\left(S_i|| \sum_j h_{ij}\mu_j\right) \\ &+ \gamma \sum_j \sum_{k\in\mathcal{N}_j} m_{jk}\phi(\mu_j - \mu_k : \delta),\end{aligned} \tag{4.13}$$

where $\phi(\mu : \delta)$ is defined in (4.11), and m_{jk} is defined as the inverse distance between neighboring pixel j and k. For image-domain regularization, δ is set to be 500. This problem is similar to maximum a-posteriori (MAP) estimation, where the penalty term serves as a prior although we have a much more complicated data fitting term.

Dual-domain regularization

A great advantage of ASAM is that we can regularize both the auxiliary sinogram domain and image domain at the same time. The introduction of variable S solves the conflict that preprocessing and the AM algorithm require different input domains. The dual-domain constrained optimization problem is formulated as

$$\begin{aligned}\min_{S,\mu} F(S,\mu) = &\min_{S,\mu} \sum_i \alpha_i I\big(d_i||Q(S_i)\big) + \beta \sum_i I\left(S_i|| \sum_j h_{ij}\mu_j\right) \\ &\text{s.t. } R(S) \le \gamma, R(\mu) \le \theta,\end{aligned} \tag{4.14}$$

and the corresponding regularized ASAM cost function is given as

$$\begin{aligned}\min_{S,\mu} F(S,\mu) = &\min_{S,\mu} \sum_i \alpha_i I\big(d_i||Q(S_i)\big) + \beta \sum_i I\bigg(S_i||\sum_j h_{ij}\mu_j\bigg) \\ &+ \gamma \sum_{i\in\tilde{C}} \sum_{k\in\mathcal{N}_i} w_{ik}\phi(S_i - S_k) \\ &+ \theta \sum_j \sum_{k\in\mathcal{N}_j} m_{jk}\phi(\mu_j - \mu_k). \end{aligned} \tag{4.15}$$

4.5.4 Minimization iterations

Optimization problem (4.15) is a dual-domain minimization. Due to the existence of parameter α and the penalty terms, we need three steps for every iteration.

Optimize $S_i, i \in C$

We fix the terms that do not contain $S_i, i \in C$, then the problem (4.15) becomes

$$\min_{S_i, i\in C} F_C(S,\mu) = \min_{S_i, i\in C} \sum_{i\in C} \alpha_i I\big(d_i||Q(S_i)\big) + \beta \sum_{i\in C} I\bigg(S_i||\sum_j h_{ij}\mu_j\bigg), \tag{4.16}$$

and it is in the same form as (4.7). However, the definition of i is limited to set C rather than the set $\{1, 2, ..., M\}$, then

$$\begin{aligned} F_C(S) &= \sum_{i \in C} \alpha_i I\big(d_i || Q(S_i)\big) + \beta \sum_{i \in C} I\bigg(S_i || \sum_j h_{ij} u_j\bigg) \\ &= \sum_{i \in C} \bigg[\alpha_i \bigg(d_i \log \frac{d_i}{Q(S_i)} - d_i + Q(S_i)\bigg) + \beta \bigg(S_i \log \frac{S_i}{\sum\limits_j h_{ij}\mu_j} - S_i + \sum_j h_{ij}\mu_j\bigg)\bigg] \\ &= \sum_{i \in C} \bigg[\alpha_i \bigg(d_i \log \frac{d_i}{I_i e^{-S_i}} - d_i + I_i e^{-S_i}\bigg) + \beta \bigg(S_i \log \frac{S_i}{\sum\limits_j h_{ij}\mu_j} - S_i + \sum_j h_{ij}\mu_j\bigg)\bigg]. \end{aligned} \tag{4.17}$$

We focus only on the terms involving S, then our problem becomes

$$F_C(S) = \sum_{i \in C} \bigg[\alpha_i \Big(- d_i S_i + I_i e^{-S_i}\Big) + \beta \Big(S_i \log S_i - S_i \log \big(\sum_j h_{ij}\mu_j\big) - S_i\Big)\bigg]. \tag{4.18}$$

Using Newton's method, $S_i, i \in C$ is updated with

$$S_i = \hat{S}_i - \bigg[\frac{\partial F_C(S_i)}{\partial S_i} \Big/ \frac{\partial^2 F_C(S_i)}{\partial S_i^2}\bigg]\bigg|_{S_i = \hat{S}_i}, \tag{4.19}$$

where $\hat{S}_i$ is the current estimate of S_i. The first order derivative for $S_i, i \in C$ is given by

$$\frac{\partial F_C(S_i)}{\partial S_i} = -d_i - I_i e^{-S_i} + \beta \bigg(\log S_i - \log \sum_j h_{ij}\mu_j\bigg), \tag{4.20}$$

where all terms that do not depend on S_i have been droped.

The corresponding second order derivative is given by

$$\frac{\partial^2 F_C(S_i)}{\partial S_i^2} = I_i e^{-S_i} + \frac{\beta}{S_i}. \tag{4.21}$$

Optimize $S_i, i \in \tilde{C}$

If we use dual-domain regularization, we take out the terms involving $S_i, i \in \tilde{C}$ from (4.15), then our problem is simplified to

$$\min_{S_i, i\in\tilde{C}} F_{\tilde{C}}(S,\mu) = \min_{S_i, i\in\tilde{C}} \sum_{i\in\tilde{C}} \alpha_i I\big(d_i||Q(S_i)\big) + \beta \sum_{i\in\tilde{C}} I\left(S_i|| \sum_j h_{ij}\mu_j\right) + \gamma \sum_{i\in\tilde{C}} \sum_{k\in\mathcal{N}_i} w_{ik}\phi(S_i - S_k). \tag{4.22}$$

The penalty term $\phi(S_i - S_k)$ couples the values of S_i and S_k. We use the convex decomposition lemma [10] to decouple this penalty. The resulting cost function for updating $S_i, i \in \tilde{C}$ is

$$\tilde{F}_{\tilde{C}}(S_i) = \alpha_i I\big(d_i||Q(S_i)\big) + \beta I\left(S_i|| \sum_j h_{ij}\mu_j\right) + \gamma \sum_{k\in\mathcal{N}_i} w_{ik}\phi\left(2S_i - \hat{S}_i - \hat{S}_k\right), \tag{4.23}$$

where $\hat{S}_i$ is the current estimated auxiliary sinogram value at pixel i. In this way, we decouple the cost function and we can update every single $S_i, i \in \tilde{C}$ at the same time. Using Newton's method, $S_i, i \in \tilde{C}$ is updated by

$$S_i = \hat{S}_i - \left[\frac{\partial \tilde{F}_{\tilde{C}}(S_i)}{\partial S_i} \Big/ \frac{\partial^2 \tilde{F}_{\tilde{C}}(S_i)}{\partial {S_i}^2}\right]\Bigg|_{S_i=\hat{S}_i}. \tag{4.24}$$

The corresponding first order derivative $\frac{\partial \tilde{F}_{\tilde{C}}(S_i)}{\partial S_i}$ and second order derivative $\frac{\partial^2 \tilde{F}_{\tilde{C}}(S_i)}{\partial {S_i}^2}$ are

$$\begin{aligned}\frac{\partial \tilde{F}_{\tilde{C}}(S_i)}{\partial S_i} =& - d_i - I_i e^{-S_i} + \frac{\beta}{\alpha_i}\left(\log S_i - \log \sum_j h_{ij}\mu_j \right) \\ &+ \frac{\gamma}{\alpha_i} \sum_{k \in \mathcal{N}_i} w_{ik} \frac{2S_i - \hat{S}_i - \hat{S}_k}{1 + \delta \left| 2S_i - \hat{S}_i - \hat{S}_k \right|}, \end{aligned} \tag{4.25}$$

$$\begin{aligned}\frac{\partial^2 \tilde{F}_{\tilde{C}}(S_i)}{\partial {S_i}^2} =& I_i e^{-S_i} + \frac{\beta}{\alpha_i S_i} \\ &+ \frac{\gamma}{\alpha_i} \sum_{k \in \mathcal{N}_j} w_{ik} \frac{2}{\left(1 + \delta \left| 2S_i - \hat{S}_i - \hat{S}_k \right|\right)^2}. \end{aligned} \tag{4.26}$$

Optimize μ_j

The cost function containing μ is

$$F_\mu(\mu) = \beta \sum_i I\left(S_i || \sum_j h_{ij}\mu_j \right) + \theta \sum_j \sum_{k \in \mathcal{N}_j} m_{jk} \phi(\mu_j - \mu_k) \tag{4.27}$$

μ_j and μ_k are also coupled [65], and we use the same convex decomposition lemma for $S_i, i \in \tilde{C}$ to decouple μ. But here, the coupling is not only in the penalty term but also in the data fitting term, so the resulting decoupled cost function for updating μ_j is

$$\begin{aligned}\tilde{F}_\mu(\mu_j) =& \beta \sum_i h_{ij}\mu_j - \beta \sum_i S_i \sum_j Z_{ij} \log \left(\left(\sum_{\hat{j}} h_{i\hat{j}} \right) \mu_j \right) \\ &+ \theta \sum_{k \in \mathcal{N}_j} m_{jk} \phi\left(2\mu_j - \hat{\mu}_j - \hat{\mu}_k \right), \end{aligned} \tag{4.28}$$

where

$$Z_{ij} = \frac{h_{ij}}{\sum_j h_{ij}}.$$

Z_{ij} is picked based on the criteria $Z_{ij} \geq 0$ and $\sum_j Z_{ij} \leq 1, \forall i$ which guarantees the convergence.

Now, we have a decoupled cost function of μ in both the data fitting term and the penalty term, and since this new cost function is still convex with respect to μ, we can compute the corresponding first order and second order derivatives

$$\frac{\partial \tilde{F}_\mu(\mu_j)}{\partial \mu_j} = \beta\Big(\sum_i h_{ij} - \sum_i Z_{ij}\frac{S_i}{\mu_j} + \Big) + \theta \sum_{k\in\mathcal{N}_j} m_{jk}\frac{2\mu_j - \hat{\mu}_j - \hat{\mu}_k}{1+\delta\left|2\mu_j - \hat{\mu}_j - \hat{\mu}_k\right|}, \tag{4.29}$$

$$\frac{\partial^2 \tilde{F}_\mu(\mu_j)}{\partial {\mu_j}^2} = \frac{\beta}{\mu_j}\sum_i Z_{ij}\frac{S_i}{\mu_j^2} + \theta \sum_{k\in\mathcal{N}_j} m_{jk}\frac{2}{\left(1+\delta\left|2\mu_j - \hat{\mu}_j - \hat{\mu}_k\right|\right)^2}. \tag{4.30}$$

With (4.29) and (4.30),using Newton's method, $S_i, i \in \tilde{C}$ is updated with

$$\mu_j = \hat{\mu}_j - \left[\frac{\partial \tilde{F}_\mu(\mu_j)}{\partial \mu_j} / \frac{\partial^2 \tilde{F}_\mu(\mu_j)}{\partial {\mu_j}^2}\right]\Bigg|_{\mu_j=\hat{\mu}_j}. \tag{4.31}$$

The overall pseudo-code for the regularized ASAM algorithm is presented in Algorithm 3.

Algorithm 3: Regularized ASAM

Initialize $S^{(0)}, \mu^{(0)}$
for $n = 0$ to $N-1$ **do**
  foreach $i \in C$ **do**
    Update $S_i^{(n+1)}, i \in C$ with (4.19), where $\hat{S}_i = S_i^{(n)}, i \in C$ and $\mu = \mu^{(n)}$.
  end
  foreach $i \in \tilde{C}$ **do**
    Fix $\hat{S}_i = S_i^{(n+1)}, i \in C$, $\mu = \mu^{(n)}$.
    Update $S_i^{(n+1)}, i \in \tilde{C}$ with (4.24).
  end
  foreach j **do**
    Update $\mu_j^{(n+1)}$ with (4.31), where $\hat{\mu}_j = \mu_j^{(n)}$ and $S_i = S_i^{(n+1)}, i \in C \cup \tilde{C}$.
  end
end

4.6 Patient phantom simulation

4.6.1 Simulation data generation

A virtual patient with metal inserts was simulated to evaluate the potential performance of ASAM. The patient phantom is a pelvis slice from the ICRP adult female phantom [24] as shown in Figure 4.4(a). Two titanium endoprostheses were inserted in the bone region. The proprietary geometry of the Philips Brilliance Big Bore CT scanner (Philips Medical Systems, Cleveland, OH) was used in this simulation. The synthetic sinogram was generated for 90 kVp and 400 mAs. In this simulation, parameters are set to be $\alpha = 0.01,\ \beta = 500,\ \gamma = 3 \times 10^4,\ \theta = 10^4$.

4.6.2 Results

The reconstructed images from FBP, NMAR, and ASAM are presented in Figures 4.4(b), (c), and (d), respectively. In the standard FBP result, there exist bright and dark streaks from the titanium implants as well as an obvious dark band connecting these two implants. In Figure 4.4(c), the streaks are successfully removed with NMAR, but the shallow band between the two implants still remains. In Figure 4.4(d), ASAM eliminates the streak artifacts and the dark band from the image. Compared with the ground truth image, NMAR and ASAM both reduce metal artifacts without sacrificing image quality.

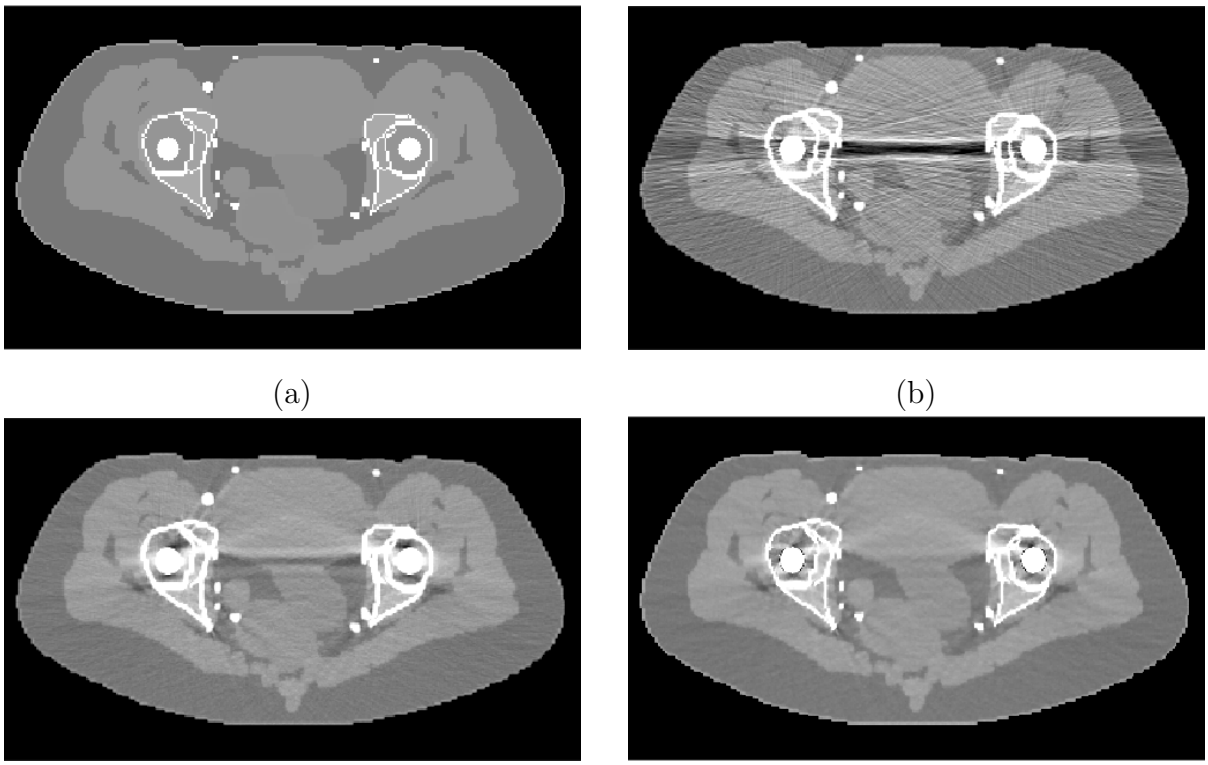


Figure 4.4: ICRP female adult pelvis phantom with titanium endoprosthesis. The ground truth is presented in (a). (b) shows the FBP result without correction. (c) is the NMAR result and (d) is the reconstructed image from ASAM. The viewing windows are [-300, 400] HU.

4.7 Physical phantom experiment

4.7.1 Experimental setup

In this section, the data of a cylinder phantom with a diameter of 21.5 cm were collected at 90 kVp (300 mAs) on the Philips Brilliance Big Bore CT scanner in the axial mode with 816 detectors per detector row and 1320 source positions per rotation. The data were collected by Dr. Dong Han at Virginia Commonwealth University. This physical phantom is an acrylic cylinder that is filled with water and contains 12 bone and soft-tissue surrogates.

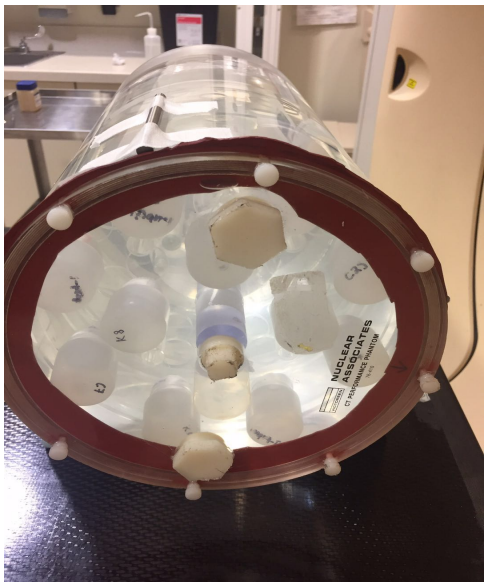
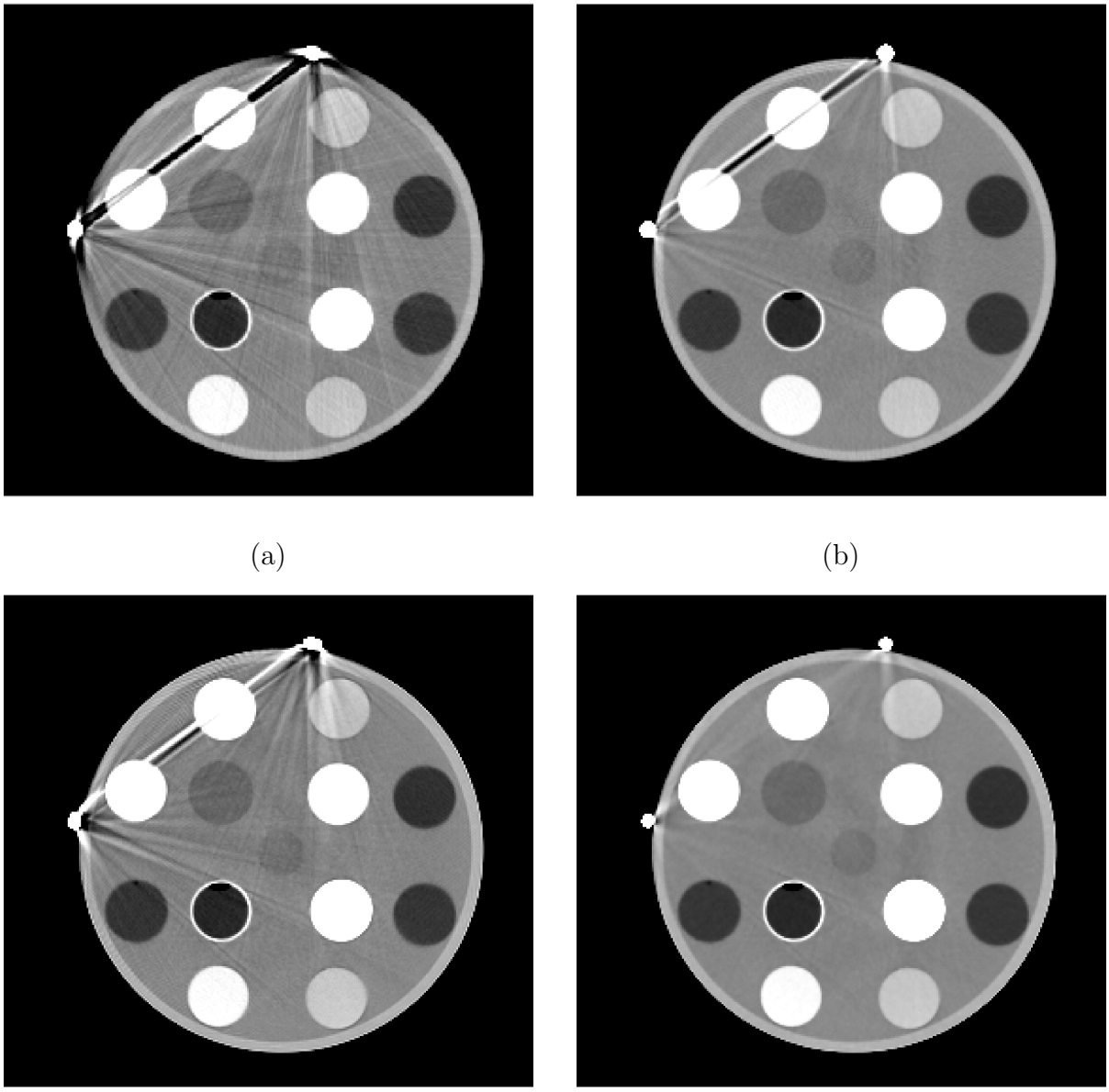


Figure 4.5: Physical phantom, and two tungsten rods adhered to this phantom.

Two tungsten rods are attached to the outside of the phantom. The reconstructed image is with pixel size 1 mm $\times$ 1 mm and slice thickness 3 mm. The reconstructed images are presented in Figure 4.6.

The standard FBP reconstruction is shown in Figure 4.6(a). We can see two major artifacts: the first is the dark streaks in the shade between two tungsten rods, and the second is streaks originating from the tungsten rods. NMAR [50] was applied and the result is shown in Figure 4.6(b). A clear decrease in streak artifacts is achieved, but the dark streaks between the rods remain. A possible reason for this underperformance is that the attenuation coefficient of tungsten is so high that no proper prior image can be defined. Then, in Figure 4.6(c), the result of the regularized alternating minimization (AM) with penalty weight $\lambda = 3 \times 10^4$ is shown. The dark streaks around the tungsten rods are reduced, but there is no significant improvement for the streaks between them. Finally, in Figure 4.6(d), we applied the ASAM with parameters $\alpha_0 = 0.01$, $\beta = 500$, $\gamma = 3 \times 10^4$, and $\theta = 10^4$. Both the shading effect and

streaking artifacts are substantially eliminated, and the tungsten rod boundaries are clearly shown.



(a) (b)

(c) (d)

Figure 4.6: Physical phantom data reconstructions from (a) filtered backprojection (FBP); (b) normalized metal-artifact reduction (NMAR); (c) alternating minimization (AM) with penalty weight $\lambda = 3 \times 10^4$; and (d) ASAM with $\alpha_0 = 0.01$, $\beta = 500$, $\gamma = 3 \times 10^4$, and $\theta = 10^4$. The viewing windows are [-300, 400] HU.

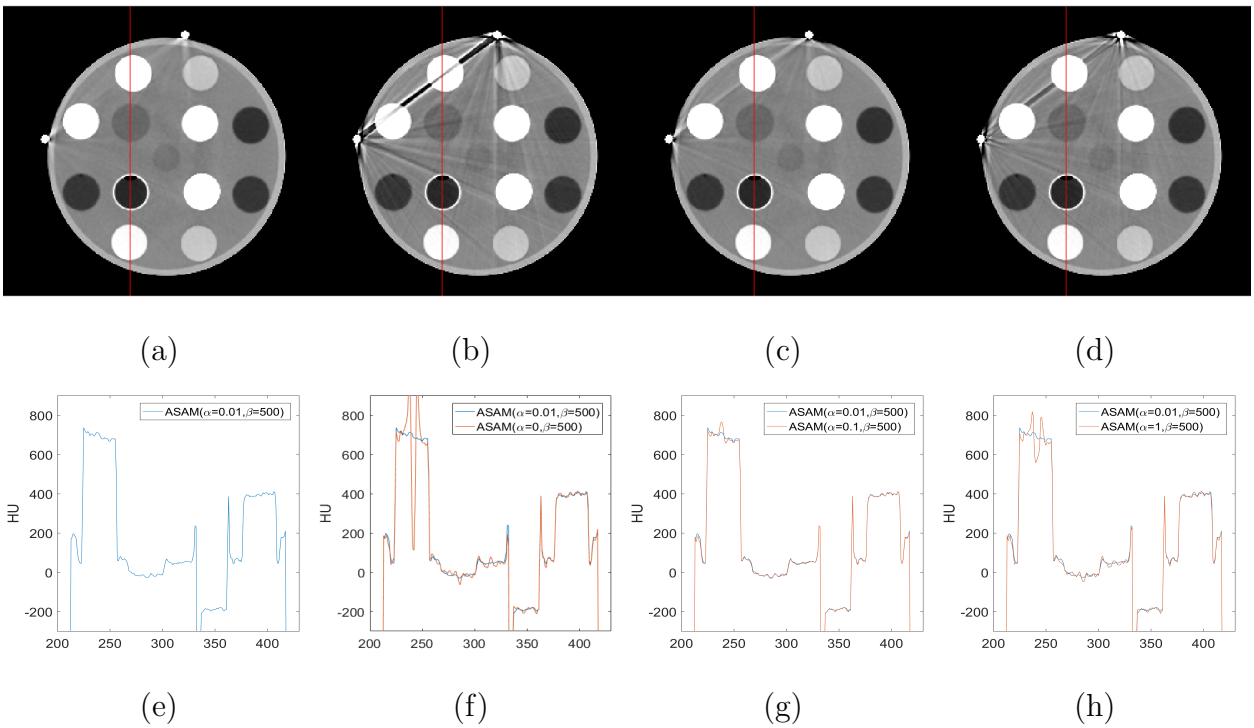


Figure 4.7: ASAM reconstructions and profiles with different α_0.

4.7.2 Results and discussion

Image-domain reconstruction

From the results in the previous section, we have proven that the ASAM algorithm is a promising method to reduce different forms of metal artifacts. A convenient property of ASAM is that for different phantoms, we only need to tune a minimal number of parameters to get the best metal-artifact reduction. The two most important parameters are α_0 and β.

α_0 controls how much the measurement affected by the metal is emphasized. Two extreme choices of α_0 are 0 and 1. $\alpha_0 = 0$ means that the metal traces are completely ignored, and $\alpha_0 = 1$ means no down-weighting is applied to the metal traces. In Figure 4.7, we compare the ASAM reconstructions with different α_0. In Figure 4.7(a), (b), (c), and (d),

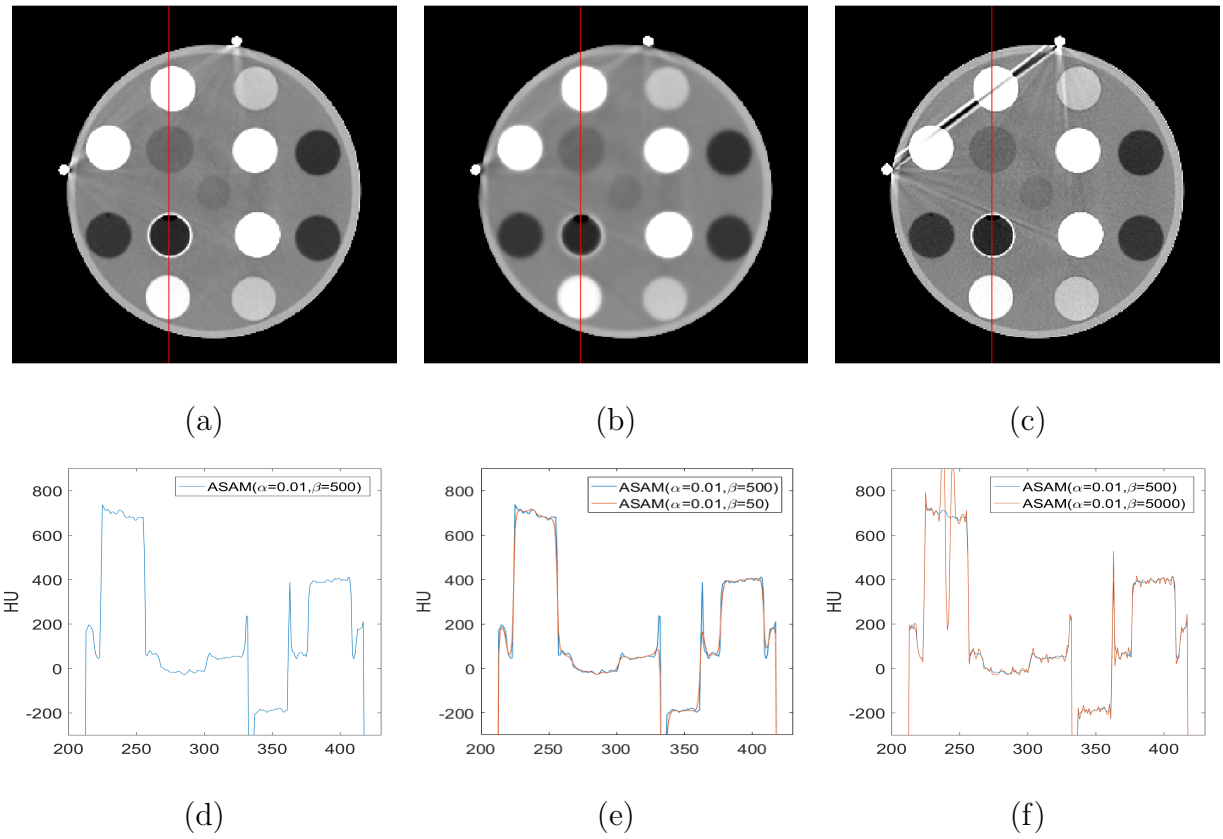
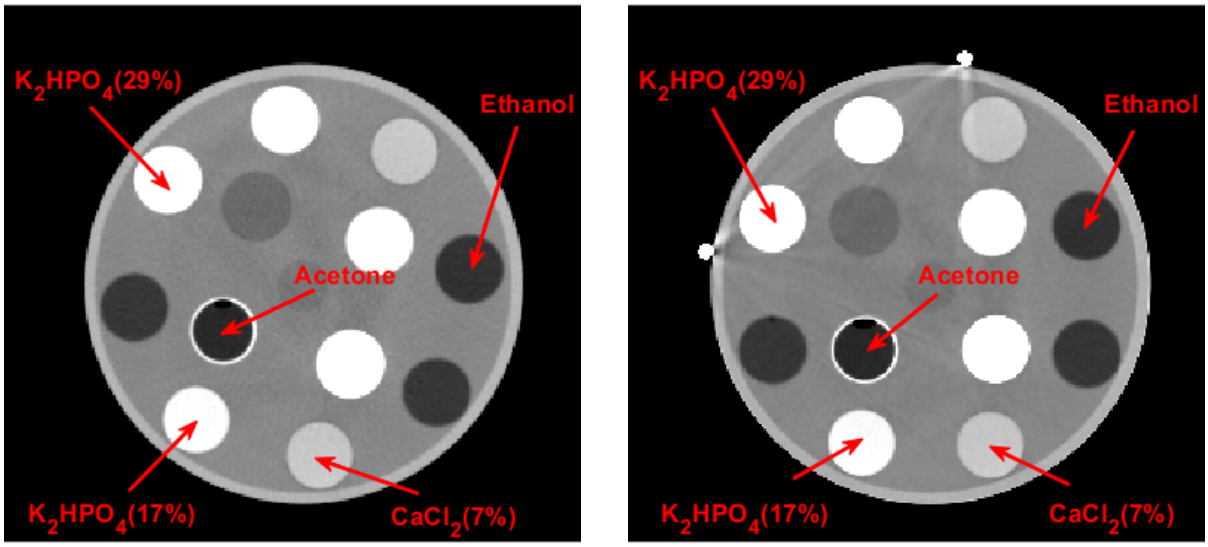


Figure 4.8: ASAM reconstructions and profiles with different β.

the ASAM reconstructions with $\alpha_0 = 0.01$, 0, 0.1, and 1 are shown. All of the results are reconstructed with $\beta = 500$, $\gamma = 3 \times 10^4$, and $\theta = 10^4$. In Figure 4.7(e), (f), (g), and (h) the corresponding profiles of the image slices highlighted by the red lines are shown. When $\alpha_0 = 0$, the reconstructed image is very close to the FBP result. The only minor image quality improvement comes from the image-domain regularization. When α_0 increases to 0.1, the metal artifacts begin to surface, and when α_0 increases to 1, which means there is no down-weight on the measurement affected by the tungsten rods, the metal artifacts become significant, and the result is very similar to the AM result.

β is the parameter that controls the tolerance of the difference between the auxiliary sinogram and the measurement sinogram. In Figure 4.8, we compare the ASAM reconstructions and



(a) (b)

Figure 4.9: (a) is the tungsten free reference image, (b) is the ASAM reconstruction with tungsten rods. The viewing window is [-200, 200] HU.

their single slice profiles with different β. In Figure 4.8(a), the metal artifacts are eliminated with $\beta = 500$. In (b), when β decreases to 50, although the metal artifacts are reduced, some resolution is lost and in the profile plot (e), the sample container edges are smoothed out. In this case, a small β means the auxiliary sinogram is distorted too much from the measurement sinogram. In Figure 4.8(c), β increases to 5000, and the metal artifacts between the two tungsten rods are not corrected. Also, the corresponding profile plot (f) shows significantly high noise and highly biased pixels brought on by the metal artifacts. A high β means a little discrepancy between the auxiliary sinogram and measurement sinogram is allowed, and the result is that little correction is done on the auxiliary sinogram domain which leads to the retention of the metal artifacts.

A concern about the ASAM algorithm is that after the measurements affected by the metal are down-weighted, the reconstruction image will suffer loss of information which may lead to bias or higher noise. To evaluate the estimation accuracy of ASAM, we reconstruct the

Samples	Reference	FBP	AM	ASAM
K_2HPO_4(29%)	521.585	473.852	486.760	510.045
K_2HPO_4(17%)	269.117	262.366	263.358	263.442
Ethanol	-244.779	-246.500	-246.370	-244.976
Acetone	-264.907	-265.642	-272.439	-267.788
$CaCl_2$(7%)	135.767	120.572	133.046	128.823

Table 4.1: Mean value of 3 sample ROIs (HU)

.

same phantom scanned without tungsten rods with the AM algorithm, which has proven to have good estimation accuracy [4], and use the reconstructed image as the reference as shown in Figure 4.9(a). In Figure 4.9(a), the reference image is free of metal artifacts but has an angular shift due to phantom rotation during the experiment. Five samples are selected: K_2HPO_4 aqueous solution (29 % by mass), K_2HPO_4 aqueous solution (17 % by mass), pure ethanol, pure acetone, $CaCl_2$ aqueous solution (7 % by mass). For each sample, the ROI is defined as a disk with radius 12 mm at the sample center. The mean values in HU of the five sample ROIs are shown in Table 4.1. For K_2HPO_4(29%), which suffers the heaviest metal artifacts, the ASAM has significant improvement in bias compared with the FBP and the AM result. For the other samples that are less affected by the tungsten rods, the ASAM has similar performance as the AM, which means that the ASAM is also an algorithm with high estimation accuracy.

Another concern about the ASAM is the convergence speed. A comparison of time to converge between the ASAM algorithm and the AM algorithm is plotted in Figure 4.10. The asymptotic cost-function values are different because AM and ASAM have different cost-function expressions. However, it is clear that ASAM has a faster decrease in the cost-function value and nearly reaches its asymptotic limit far earlier than the AM algorithm. This rapid convergence is because instead of enforcing the equality between the reconstructed

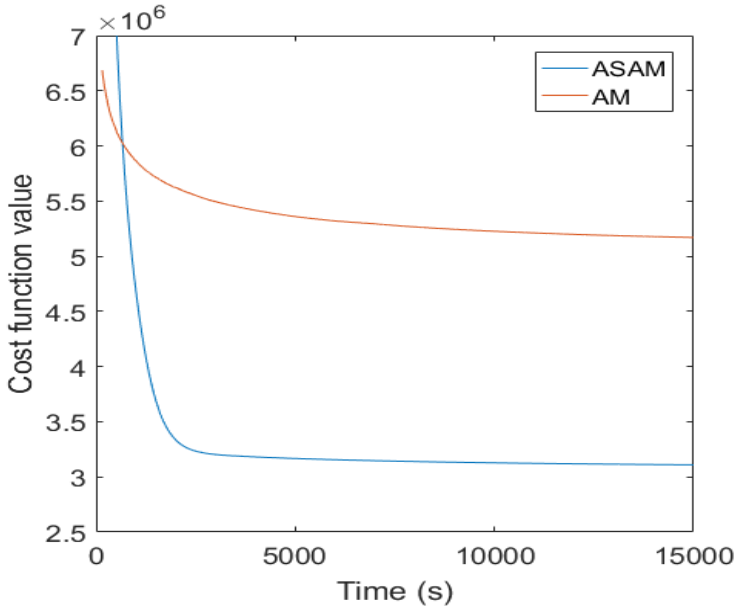


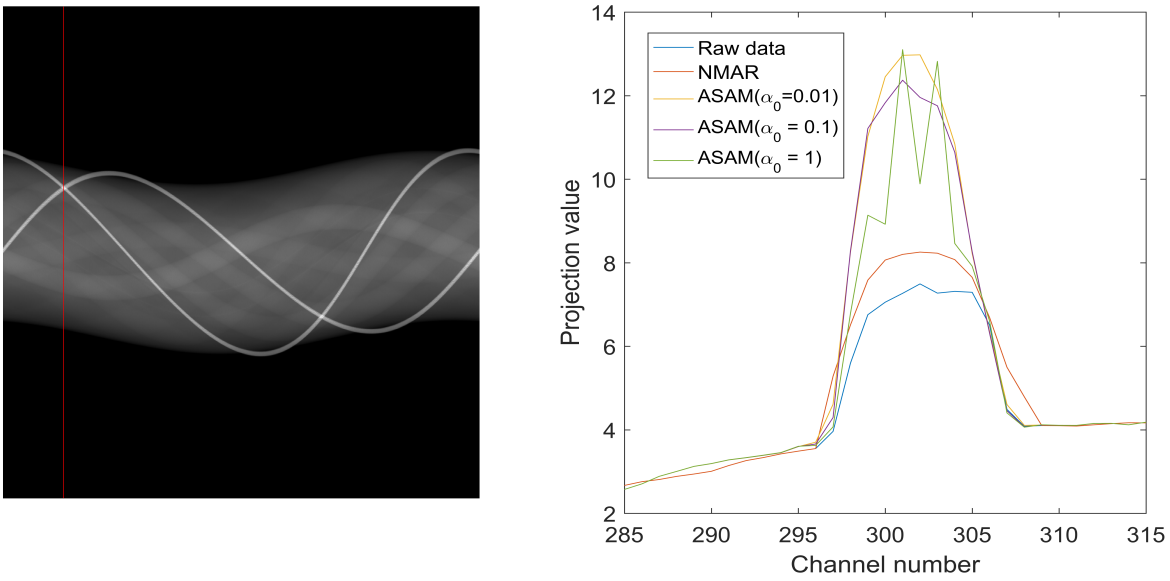
Figure 4.10: The time required by AM and ASAM to converge.

auxiliary sinogram and the forward projection of the reconstructed image, a discrepancy is allowed in the form of the term $\beta \sum_i I(S_i || \sum_j h_{ij}\mu_j)$ in the cost function [23]. As for the AM algorithm, the only time-consuming part in the ASAM algorithm is the forward- and backward-projection computation. In this chapter, 33, 11, and 5 ordered subsets [66, 67] were used sequentially to speed up the computation.

Sinogram-domain reconstruction

ASAM is a dual-domain algorithm. Besides the image-domain output μ, the auxiliary sinogram S is another output. By studying the auxiliary sinogram reconstruction, we can further analyze the performance of the ASAM algorithm.

The reconstructed auxiliary sinogram S is plotted in Figure 4.11(a). The slice where two tungsten rods align in the same projection angle is selected as highlighted with the red



(a) (b)

Figure 4.11: (a) the reconstructed auxiliary sinogram, (b) the profile of a single sinogram slice highlighted with a red line from channel 285 to channel 315.

line. The sinogram profile of this slice from channel 285 to channel 315, where the peak value corresponds to where the tungsten rods are placed, is shown in Figure 4.11(b). Due to the circular shape of the rods and the high attenuation of tungsten, the ground truth shape of this peak value should be a sharp arc. However, the raw attenuation sinogram, which is the blue curve, shows a flat top. This is because when the attenuation is high in these channels, the scatter-to-primary ratio (SPR) dramatically increases, which results in overestimation of the primary photons. When the SPR is high, the attenuation is underestimated, and the scatter causes the flat top in the raw sinogram. In NMAR, the tungsten pixel values of the prior image are set from the FBP result, which is underestimated due to the scatter. We can see that the arc shape is recovered with the NMAR method, but the underestimation is not compensated as the red curve shows. With ASAM, a proper choice of α_0 to down-weight the metal projections can compensate the underestimation caused by

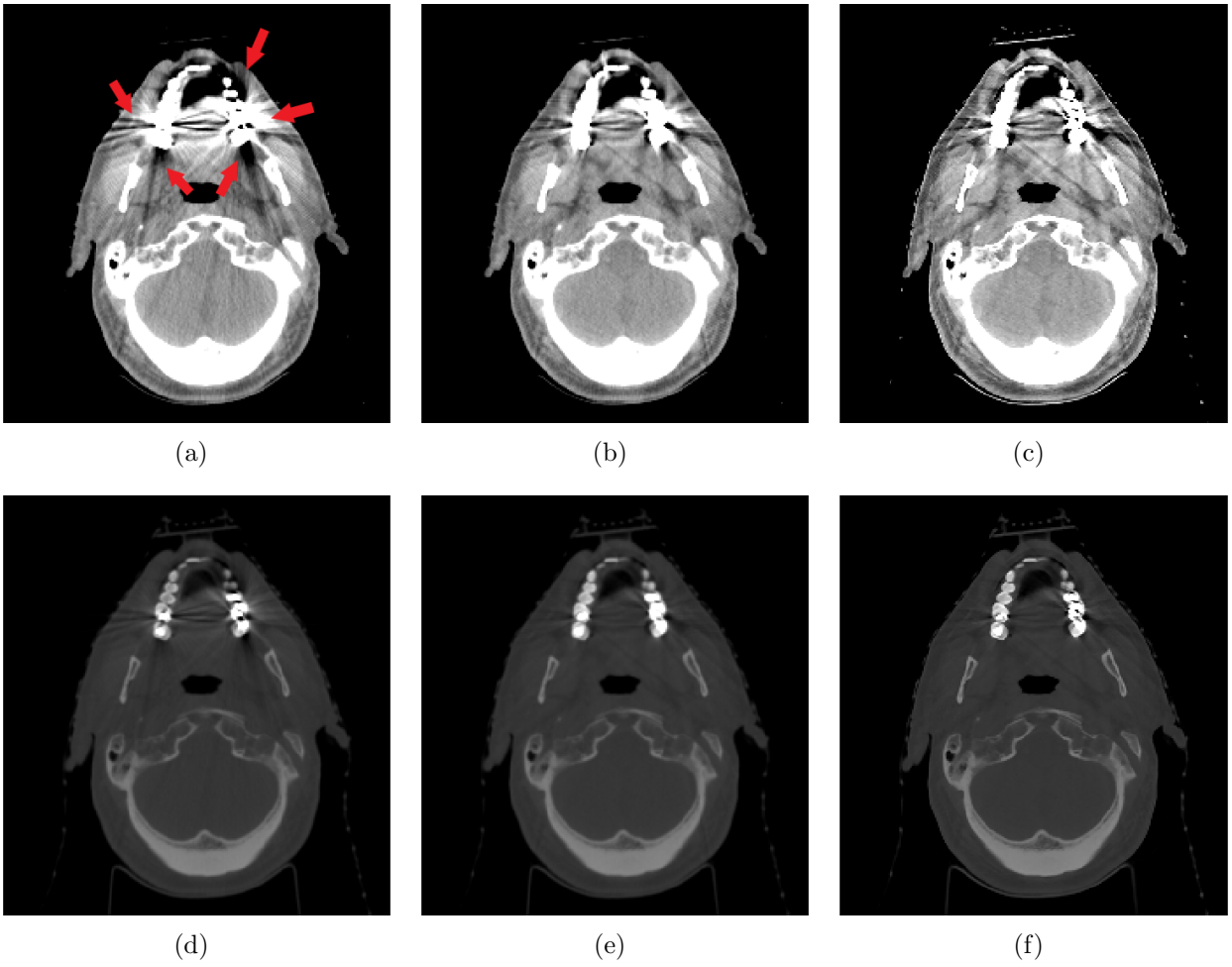
scatter and recover the sharp arc shape as the yellow and purple curves show. When there is no down-weighting, which means $\alpha_0 = 1$, the curve is underestimated and noisy. This is because in function (4.22), the effective sinogram-domain penalty weight is $\frac{\gamma}{\alpha_0}$, and $\alpha_0 = 1$ results in a low effective penalty weight compared with other small α_0 choices. Also, this sinogram reconstruction result demonstrates that the sinogram-domain penalty can serve as the interpolation operator that approximately recovers the ground truth sinogram shape.

Although the NMAR method can recover the arc shape, the underestimation results in the dark shade between two tungsten rods, which is the significant difference between the ASAM reconstruction and the NMAR reconstruction. The performance of NMAR heavily relies on the choice of the prior image. In the experiment, the prior tungsten pixels from the FBP reconstruction cannot compensate for the underestimation due to scatter. ASAM can be viewed as an iterative version of NMAR. Instead of using an initial constant prior image, the "prior image" is updated at every iteration in ASAM. Combined with the choice of α_i, the underestimation can be compensated for in ASAM.

4.8 Patient scan results

A patient was scanned on the Philips Brilliance Big Bore CT scanner (140 kVp, 300 mAs, 12 mm collimation) in the axial mode with 816 detectors per detector row and 1320 source positions per rotation. The pixel size of the reconstructed image is 1 mm $\times$ 1 mm.

The patient has multiple dental fillings, which are made of high-density materials and have irregular shapes. Compared with other tissues, dental enamel is the densest material in the human body. Metal-artifact reduction, in this case, is very challenging.



(a) (b) (c)

(d) (e) (f)

Figure 4.12: Patient head-scan reconstructions: (a) FBP, (b) NMAR, (c) ASAM with $\alpha_0 = 0.1$, $\beta = 500$, $\gamma = 3 \times 10^4$, $\theta = 10^4$. The viewing windows are [-200, 200] HU. (d), e, and (f) are the corresponding reconstructions with viewing windows [-1000, 3500] HU.

Fig. 4.12 shows a slice through the jaw with multiple dental fillings. In Fig. 4.12(a), with a narrow viewing window, the image generated with the FBP algorithm suffers heavy metal artifacts. The tooth boundaries can hardly be identified. With a large viewing window as shown in Fig. 4.12(d), the artifacts are along the direction where the dental fillings are placed.

Both NMAR and ASAM can effectively reduce metal artifacts, especially the dark shades and streaks in the back of the teeth, and give a much better boundary identification of the tooth boundaries. However, there are residual artifacts for both methods. This is because the irregular shapes of the dental fillings make it difficult to define an optimal prior image in the NMAR method and also the optimal definition of the set $\tilde{C}$ in the ASAM algorithm.

4.9 Summary

We introduced the new ASAM algorithm that uses an auxiliary sinogram domain to connect raw measurement data and the reconstructed image. This auxiliary sinogram domain gives us the flexibility to include prior knowledge which was previously available in the image domain. By solving this dual-domain optimization problem, we can reconstruct significantly more accurate and artifact-free images. Preprocessing methods and iterative reconstruction algorithms are both powerful tools in reducing metal artifacts. The new ASAM algorithm can integrate these two categories and avoid their incompatibilities that block the idea of combining them. Their common drawback, that they ignore the metal traces data, which leads to loss of information, is solved by the introduction of an auxiliary sinogram. We use the parameter α_0 to suppress the artifacts from these metal traces. By the choices of parameters and the forms of penalties, we have the freedom to apply different interpolation methods or even new preprocessing methods in the same scheme. In other words, our ASAM is a general solution to preprocessing methods.

Chapter 5

Wavelet alternating minimization algorithm (wav-AM) for x-ray CT

5.1 Highlights

In the previous chapters, it was shown that the neighborhood penalties in different data domains successfully reduce noise and suppress artifacts. However, in some x-ray CT applications, the performance of the traditional neighborhood penalty is limited, and a transform-domain penalty is used. In this chapter, a wavelet domain penalty is introduced for security x-ray scans to help reduce streak artifacts.

5.2 Motivation

In the previous chapters, it was shown that the quality of the reconstructed image is affected by many factors. For example, if there exist dense or metal materials in the scanned object,

the image suffers heavy artifacts. If the number of detected photons is low, reconstructed images have high noise. One of the widely used methods to reduce noise and suppress artifacts is to use the statistical image reconstruction algorithm which is a regularized optimization method. The penalty term contains the prior information about the image, which usually promotes the smoothness of the image. Typical choices of the penalty term are total variation or Huber-type penalties [68, 12, 69]. The choice of the penalty weights controls the balance between the data fitting term and the penalty term. A low penalty weight retains the resolution but does not suppress the noise or artifacts. A high penalty weight can accomplish the goal of noise reduction and artifact suppression, but the trade-off is a possible loss of detail in resolution, or over-smoothness. For low-dose x-ray CT scans, choosing a good penalty weight is not only critical but also difficult. The contradictory requirements favoring a low weight to improve some image characteristics and a high weight to emphasize other characteristics make it difficult to find an optimal choice. To solve this problem, another category of prior information of sparsity in the wavelet coefficient domain is used. Based on this transform-domain penalty, the wavelet alternating minimization algorithm is developed, which has the following properties:

- The data fitting term is still maximum Poisson log-likelihood estimation with the mean determined by Beer's law.
- The sparsity penalty in the wavelet domain is easy to formulate and easy to implement.
- The wav-AM algorithm can be extended to a dual-domain regularization problem where another image-domain penalty is added.

5.3 Notation

In this section, the notations used in this chapter are summarized.

- d: measured transmission data.
- y: estimated mean.
- u: reconstructed image.
- β: wavelet transform of the reconstructed image
- I: air-scan source intensity.
- H: system matrix.
- Ω: inverse wavelet transform matrix.
- $\hat{\Omega}$: wavelet transform matrix.
- i: source-detector pair index.
- j: image pixel (voxel) index.
- l: wavelet coefficient index.

5.4 Image-domain regularization

In the previous chapters, it was shown that minimizing the I-divergence between the measured data and the estimated mean is equivalent to maximizing the Poisson log-likelihood,

and the problem is written as

$$\min_u \sum_i I(d_i||y_i), \tag{5.1}$$

where

$$y_i = I_i e^{-\sum_j h_{ij} u_j} \tag{5.2}$$

is the estimated mean following Beer's law [70]. Because there exists no closed form solution for this problem, the AM algorithm is used to solve this problem iteratively by minimizing a surrogate function that approximates the original cost function. We simplify (2.22) to a single-energy model, without the penalty term, u_j is updated in a closed form

$$u_j = \hat{u}_j - \frac{1}{Z_0} \log \frac{\sum_i h_{ij} d_i}{\sum_i h_{ij} y_i}, \tag{5.3}$$

where $\hat{u}_j$ is the current reconstructed image pixel, and Z_0 is a parameter that guarantees convergence.

If an image-domain penalty is included in the cost function, there exists no closed form update. Instead, the convex decomposition lemma is used to get the surrogate function that approximates the penalty term. Then, Newton's method is used to update the reconstructed image. The penalty term contains prior information about the image, and the edge-preserving Huber-type penalty is used in the following form

$$\phi(x) = \frac{1}{\delta^2}(|\delta x| - \log(1 + |\delta x|)), \tag{5.4}$$

where δ is a parameter that controls the transition between a quadratic region (for small x) and a linear region (for large x).

The image domain regularized AM cost function is formulated as

$$f(u) = \sum_i I(d_i||y_i) + \lambda \sum_j \sum_{k \in \mathcal{N}_j} w_{ij}\phi(u_j - u_k), \tag{5.5}$$

where $\mathcal{N}_j$ is the set of neighborhood pixels or voxels of voxel j. The method for minimizing the cost function (5.5) was introduced in Chapter 2.

5.5 Wavelet alternating minimization algorithm

5.5.1 Unregularized wavelet alternating minimization

Assume there exists a discrete orthogonal wavelet transform pair $\Omega \in R^{N \times M}$ and $\hat{\Omega} \in R^{M \times N}$ such that

$$u = \Omega \times \beta \tag{5.6}$$

and

$$\beta = \hat{\Omega} \times u, \tag{5.7}$$

where $\beta \in R^M$ is the vector of wavelet coefficients. Based on this transform pair, the function (5.1) can be rewritten as

$$f(\beta) = \sum_i \Big[d_i \sum_j h_{ij} \sum_l \omega_{jl}\beta_l + I_i \exp(-\sum_j h_{ij} \sum_l \omega_{jl}\beta_l) \Big]. \tag{5.8}$$

For simplicity, a new system matrix Ψ is defined as $\Psi = H \times \Omega$, such that $\psi_{il} = \sum_j h_{ij}\omega_{jl}$. The above cost function is reformulated as

$$f(\beta) = \sum_i \Big[d_i \sum_l \psi_{il}\beta_l + I_i \exp(-\sum_l \psi_{il}\beta_l) \Big], \tag{5.9}$$

where ψ_{il} is the new system matrix element, and β is the wavelet coefficient that needs to be reconstructed. Instead of directly minimizing function (5.9), a surrogate function is used. However, extra caution should be taken that every element of H and u is non-negative. But in function (5.9) either ψ_{il} or β_l can be negative. Some modification is applied in the derivation of the surrogate function. By the convex decomposition lemma [4]

$$\begin{aligned} f(\beta) &= \sum_i \Big[d_i \sum_l \psi_{il}\beta_l + I_i \exp(-\sum_l \psi_{il}\beta_l) \Big] \\ &= \sum_i \sum_l d_i \psi_{il}\beta_l + \hat{q}_i \exp(-\sum_l \psi_{il}(\beta_l - \hat{\beta}_l)) \\ &\leq \sum_i \sum_l d_i \psi_{il}\beta_l + \sum_i \sum_l \hat{q}_i r_{il} \exp(-\frac{\psi_{il}}{r_{il}}(\beta_l - \hat{\beta}_l)), \end{aligned} \tag{5.10}$$

where $\hat{q}_i = I_i \exp(-\sum_l \psi_{il}\hat{\beta}_l)$ is the forward projection of the current estimate. r_{il} is a parameter with constraints $r_{il} \geq 0$ and $\sum_l r_{il} \leq 1$. In this chapter, r_{il} is defined as

$$r_{il} = \frac{|\psi_{il}|}{Z_0}, \tag{5.11}$$

where

$$Z_0 = \max_i \sum_l |\psi_{il}|. \tag{5.12}$$

The new surrogate function is

$$\begin{aligned}\hat{f}(\beta) &= \sum_i \sum_l d_i \psi_{il} \beta_l + \sum_i \sum_l \hat{q}_i r_{il} \exp(-\frac{\psi_{il}}{r_{il}}(\beta_l - \hat{\beta}_l)) \\ &= \sum_i \sum_l d_i \psi_{il} \beta_l + \sum_i \sum_l \hat{q}_i \frac{|\psi_{il}|}{Z_0} \exp(-\frac{\psi_{il}}{|\psi_{il}|} Z_0(\beta_l - \hat{\beta}_l)) \\ &= \sum_i \sum_l d_i \psi_{il} \beta_l + \sum_{\{i,l\} \in C^+} \hat{q}_i \frac{\psi_{il}}{Z_0} \exp(-Z_0(\beta_l - \hat{\beta}_l)) - \sum_{\{i,l\} \in C^-} \hat{q}_i \frac{\psi_{il}}{Z_0} \exp(Z_0(\beta_l - \hat{\beta}_l)),\end{aligned} \tag{5.13}$$

where $C^+ = \big\{\{i,l\} : \psi_{il} \geq 0\big\}$ and $C^- = \big\{\{i,l\} : \psi_{il} < 0\big\}$. In this way, the exponential term is separated by the positivity of ψ_{il}.

To minimize the surrogate function (5.13), the first-order derivative is computed

$$\frac{\partial \hat{f}}{\partial \beta_l} = \sum_i d_i \psi_{il} - \sum_{i \in C^+} \hat{q}_i \psi_{il} \exp(-Z_0(\beta_l - \hat{\beta}_l)) - \sum_{i \in C^-} \hat{q}_i \psi_{il} \exp(Z_0(\beta_l - \hat{\beta}_l)). \tag{5.14}$$

Let $a_l = \sum_{i \in C^-} -\hat{q}_i \psi_{il}$, $c_l = \sum_{i \in C^+} \hat{q}_i \psi_{il}$, $b_l = \sum_i d_i \psi_{il}$, $x_l = \exp(Z_0(\beta_l - \hat{\beta}_l))$. Then the above function equal to zero gives

$$b_l - \frac{c_l}{x_l} - a_l x_l = 0 \Leftrightarrow a_l x_l^2 - b_l x_l + c_l = 0, \text{if} x_l \neq 0. \tag{5.15}$$

The solution is directly given by

$$x_l = \frac{b_l \pm \sqrt{b_l^2 - 4 a_l c_l}}{2 a_l}. \tag{5.16}$$

From the fact that $x_l = \exp\left(Z_0(\beta_l - \hat{\beta}_l)\right) > 0$, $a_l < 0$, and $c_l > 0$, $b_l < \sqrt{b_l^2 - 4a_l c_l}$. The only valid solution is

$$x_l = \frac{b_l - \sqrt{b_l^2 - 4a_l c_l}}{2a_l}. \tag{5.17}$$

The closed form update is

$$\beta_l = \hat{\beta}_l + \frac{1}{Z_0} \log \frac{b_l - \sqrt{b_l^2 - 4a_l c_l}}{2a_l}. \tag{5.18}$$

The algorithm of unregularized wavelet AM is shown below.

Algorithm 4: Unregularized wav-AM algorithm

Precompute Z_0 and $q^{(0)}$. Initialize $\beta^{(0)}$.
for $n = 1$ to N **do**
  foreach i **do**
    Update $q_i^{(n)}$ with $q_i^{(n)} = I_i \exp(-\sum_l \psi_{il} \beta_l^{(n)})$.
  end
  foreach l **do**
    Compute $a_l^{(n)}$, $b_l^{(n)}$, and $c_l^{(n)}$.
    Update the estimate with $\hat{\beta}_l^{(n+1)} = \hat{\beta}_l^{(n)} + \frac{1}{Z_0} \log \frac{b_l^{(n)} - \sqrt{{b_l^{(n)}}^2 - 4a_l^{(n)} c_l^{(n)}}}{2a_l^{(n)}}$
  end
end

5.5.2 Wavelet-domain regularization

Wavelet thresholding [71] is a widely used denoising technique in image processing, and the underlining assumption is that the wavelet coefficient of the ground truth image is sparse compared to the noisy image. This assumption serves as the prior knowledge in the form of

an L1 penalty [72, 73]. The regularized wavelet alternating minimization cost function is

$$f(\beta) = \sum_i I(d_i||y_i) + \gamma \sum_l |\beta_l|. \tag{5.19}$$

Because the penalty term is self decoupled, there is no need to decouple it and the surrogate function for the regularized wav-AM is

$$\hat{f}(\beta) = \sum_i \sum_l d_i \psi_{il} \beta_l + \sum_{\{i,l\} \in C^+} \hat{q}_i \frac{\psi_{il}}{Z_0} \exp\left(-Z_0(\beta_l - \hat{\beta}_l)\right) - \sum_{\{i,l\} \in C^-} \hat{q}_i \frac{\psi_{il}}{Z_0} \exp\left(Z_0(\beta_l - \hat{\beta}_l)\right) + \gamma \sum_l |\beta_l|. \tag{5.20}$$

The corresponding first order derivative with respect to β_l is

$$\frac{\partial \hat{f}}{\partial \beta_l} = \sum_i d_i \psi_{il} - \sum_{i \in C^+} \hat{q}_i \psi_{il} \exp\left(-Z_0(\beta_l - \hat{\beta}_l)\right) - \sum_{i \in C^-} \hat{q}_i \psi_{il} \exp\left(Z_0(\beta_l - \hat{\beta}_l)\right) + \gamma \mathrm{sgn}(\beta_l). \tag{5.21}$$

Setting this first order derivative to zero gives

$$\sum_i d_i \psi_{il} - \sum_{i \in C^+} \hat{q}_i \psi_{il} \exp\left(-Z_0(\beta_l - \hat{\beta}_l)\right) - \sum_{i \in C^-} \hat{q}_i \psi_{il} \exp\left(Z_0(\beta_l - \hat{\beta}_l)\right) + \gamma \mathrm{sgn}(\beta_l) = 0. \tag{5.22}$$

The $\mathrm{sgn}(\beta)$ is removed by the positivity of β_l.

If $\beta_l \geq 0$, with the definitions of a_l, b_l, c_l, and x_l, the wavelet coefficient is updated with

$$\beta_{l+} = \max\left(\hat{\beta}_l + \frac{1}{Z_0} \log \frac{b_l - \sqrt{b_l^2 - 4a_l c_l}}{2a_l}, 0\right). \tag{5.23}$$

If $\beta_l < 0$, with the definitions of a_l, b_l, c_l, and x_l, the wavelet coefficient is updated with

$$\beta_{l-} = \min\left(\hat{\beta}_l + \frac{1}{Z_0} \log \frac{b_l - \sqrt{b_l^2 - 4a_l c_l}}{2a_l}, 0\right). \tag{5.24}$$

Compared with the image-domain regularized AM, there exists a closed form update in regularized wav-AM, which is a significant advantage in computation speed. The wavelet domain regularized wav-AM is summarized in Algorithm 5.

Algorithm 5: Regularized wav-AM algorithm

Precompute Z_0 and $q^{(0)}$. Initialize $\beta^{(0)}$.
for $n = 1$ to N **do**
  foreach i **do**
    Update $q_i^{(n)}$ with $q_i^{(n)} = I_i \exp(-\sum_l \psi_{il}\beta_l^{(n)})$.
  end
  foreach l **do**
    Compute $a_l^{(n)}$, and $c_l^{(n)}$.
    Let $b_l^{(n)} = \sum_i d_i\psi_{il} + \gamma$, update $\beta_{l+}^{(n)}$ based on (5.23).
    Let $b_l^{(n)} = \sum_i d_i\psi_{il} - \gamma$, update $\beta_{l-}^{(n)}$ based on (5.24).
    Update $\beta_l^{(n+1)} = \max\left(\beta_{l+}^{(n)}, \min(\beta_{l-}^{(n)}, 0)\right)$.
  end
end

5.5.3 Phantom simulation

In this subsection, the performance of the proposed wav-AM algorithm for x-ray CT is examined with the simulated data. Four methods are compared: (1) the proposed unregularized wav-AM algorithm; (2) the regularized wav-AM algorithm with different choices of the penalty weight γ; (3) the unregularized AM algorithm; and (4) the regularized AM algorithm with different choices of the penalty weight λ.

The simulated data were generated from a FORBILD phantom [74] with Poisson noise, and the mean is generated with transmission according to Beer's law. The dose level is set to be 150 mAs which is typical for clinical systems. The image has size 256×256 pixels with pixel

size 1 mm $\times$ 1 mm. The number of source-detector pairs is 21600. The wavelet transform is chosen to be a Haar wavelet with 3 levels.

The reconstructed images are shown in Figure 5.1 and the detail structures in the red boxes are shown in the upper-left corners. In Figure 5.1(a), the ground truth image is presented. In Figure 5.1(b), the reconstructed image with the AM algorithm is shown, and since the Haar wavelet transform with 3 levels is an orthogonal operator, it is also the reconstructed image with the unregularized wav-AM algorithm. Clearly, noise can be seen in this image. In Figures 5.1(c) - (e), the reconstructed images with regularized AM with penalty weights 2×10^4, 2×10^5, and 6×10^5 are shown. In Figure 5.1(c), the penalty weight is small and the noise is still prominent. In Figure 5.1(d), the penalty weight choice is appropriate and the detail resolution and noise reduction are kept at the same time. In Figure 5.1(e), the penalty weight is increased to 6×10^5, The image is over-smoothed, and the detail structures are blurred. In Figure 5.1(f) - (h), the reconstructed images with the wavelet regularized AM algorithm are presented. In Figure 5.1(f), the penalty weight is 500, and the noise level does not get sufficiently suppressed. In Figure 5.1(g), the penalty weight choice is set to be 1000, and we have a good balance between smoothness and the detail resolution. If the penalty weights increases to 2000 as in Figure 5.1(h), the detail structure resolution is lost and a blocky effect on the edges can be seen.

To further compare the performance, the profiles of slice No. 199 which are highlighted with red lines in Figure 5.1 are presented in Figure 5.2. Figure 5.2(a) shows the ground truth profile. In Figure 5.2(b), the unregularized AM/wav-AM reconstruction result is shown, and severe noise can be seen. From Figures 5.2(c) to (e), with increased penalty weight, the noise level is reduced. However, the peak and valley contrast is also lost. The peak and valley values shrink towards their average value. The smoothness is improved at the cost

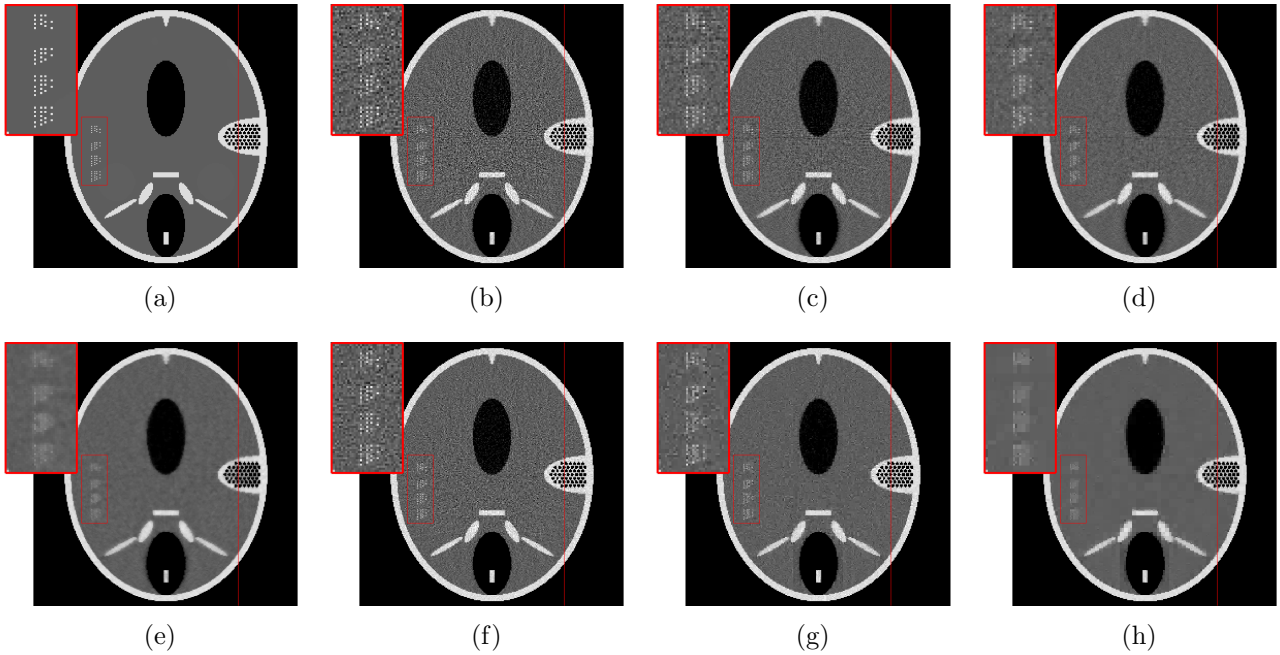


Figure 5.1: Simulated data reconstructions: (a) the ground truth, (b) unregularized AM/wav-AM, (c) regularized AM with $\lambda = 2 \times 10^4$, (d) regularized AM with $\lambda = 2 \times 10^5$, (e) regularized AM with $\lambda = 6 \times 10^5$, (f) regularized wav-AM with $\gamma = 500$, (g) regularized wav-AM with $\gamma = 1000$, and (h) regularized wav-AM with $\gamma = 2000$.

of peak and valley contrast resolution[75]. From Figures 5.2(f) to (h), with the increase of the wavelet penalty weight γ from 500 to 2000, the noise level in the homogeneous region is reduced, the peak and valley contrast is kept.

From this simulation, the wavelet penalty has shown a significant advantage over the traditional image-domain neighborhood penalty. With a large penalty weight in the neighborhood penalty, the improvement in the smoothness is at the sacrifice of detail resolution and edge contrast. The wavelet penalty can retain the detail resolution and edge contrast.

Quantitatively, the root mean square error (RMSE) and the peak signal-to-noise ratio (PSNR) are compared and shown in Table 5.1. All these regularized algorithms improve performance compared with the unregularized AM/wav-AM method. With the best choices

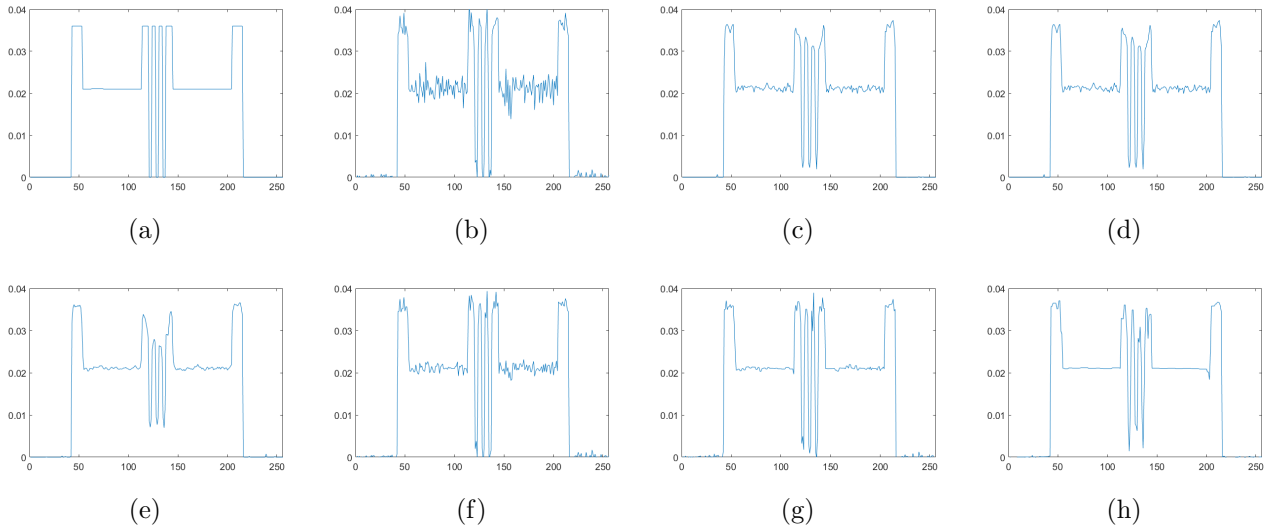


Figure 5.2: Profiles of slice no. 199: (a) the ground truth, (b) unregularized AM/wav-AM, (c) regularized AM with $\lambda = 2 \times 10^4$, (d) regularized AM with $\lambda = 2 \times 10^5$, (e) regularized AM with $\lambda = 6 \times 10^5$, (f) regularized wav-AM with $\gamma = 500$, (g) regularized wav-AM with $\gamma = 1000$, and (h) regularized wav-AM with $\gamma = 2000$.

	AM/wav-AM	Wavelet regularized AM			Neighborhood regularized AM		
		$\gamma = 500$	$\gamma = 1000$	$\gamma = 2000$	$\lambda = 2 \times 10^4$	$\lambda = 2 \times 10^5$	$\lambda = 6 \times 10^5$
RMSE	20.29×10^{-4}	7.48×10^{-4}	6.95×10^{-4}	10.26×10^{-4}	$12.18 \times 10{-}4$	10.45×10^{-4}	12.10×10^{-4}
PSNR(dB)	26.74	34.79	35.72	32.15	30.95	31.79	29.88

Table 5.1: RMSE using traditional AM with different penalty weights and wav-AM using different penalty weights.

in penalty weights, $\gamma = 1000$ and $\lambda = 2 \times 10^5$, the wavelet regularized AM outperforms the traditional neighborhood regularized AM algorithm in both RMSE and PSNR, which is in agreement with the observation in Figure 5.2.

5.5.4 Bag data experiment

In the simulated data experiment, although both the regularized AM and regularized wav-AM can reduce the noise, an optimal choice of the penalty weight offers the best performance.

However, in the following bad data experiment, it is shown that it is difficult to find an optimal weight.

Data package were acquired from a SureScan^{TM} $x1000$ explosive detection system which scanned a NIST phantom inside a suitcase. The object of interest was a Delrin cylinder wrapped with aluminum, copper, tin and lead. The geometry is shown in Figure 5.3 [20].

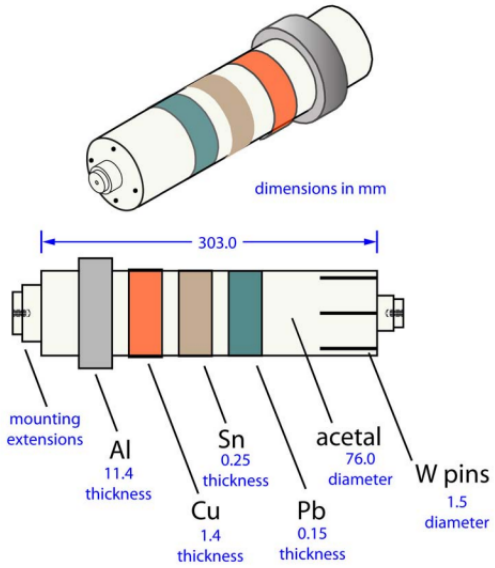


Figure 5.3: Geometry of the NIST A phantom. Figure courtesy of SureScan^{TM} Corporation.

In Figure 5.4, the histogram of the source-detector pairs with respect to the number of detected photons is plotted. It can be seen that the maximum number of detected photons for any single source-detector pair is fewer than 2000, and a large portion of the source-detector pairs have a minimal number of photon counts. The data is photon starved and suffers heavy noise and artifacts. The methods introduced above were used to reconstruct the 3-D volume.

In Figure 5.5, a lateral slice of the reconstructed volume with the regularized AM and regularized wav-AM algorithms is illustrated. In Figure 5.5(a), the reconstruction with the

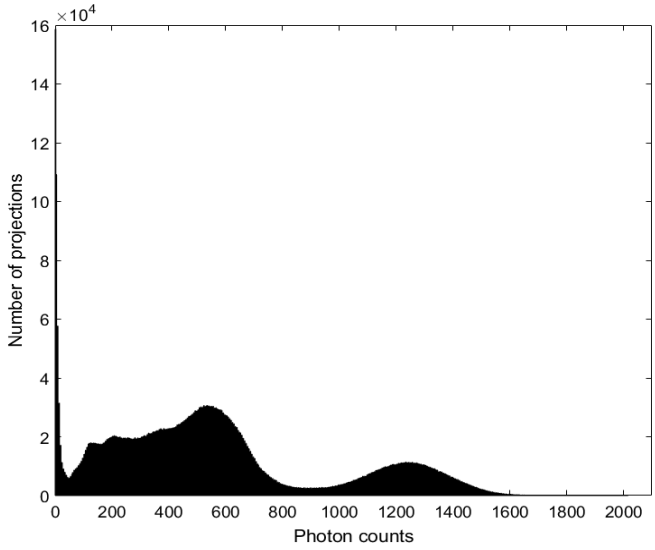


Figure 5.4: Histogram of photon counts at source-detector pairs.

unregularized AM algorithm is shown. Prominent noise is present across the Derlin cylinder. In Figure 5.5(b), (c), and (d), the reconstructed images with the regularized AM algorithm with penalty weights $\lambda = 1000$, 5000, and 15000 are shown. With the increase of λ, the noise is suppressed, but also over-smoothed. Figure 5.5(e) and (f) show the reconstructed images with the regularized wav-AM and $\gamma = 1000$ and 3000. With the increase of γ, the noise is also suppressed and the smoothness is improved. However, even with large γ, the smoothness with wav-AM is not as good as the traditional regularized AM with a large penalty weight λ.

Although with a large $\lambda = 15000$, smooth images can be reconstructed, there exist disadvantages because in some scenarios, the penalty weight cannot be allowed to be too large. In Figure 5.6, the profiles of a single column are illustrated. The column is denoted as the red line in Figure 5.5(a). In Figure 5.6(a), the profiles of the unregularized AM ($\lambda = 0$) and regularized AM with $\lambda = 1000$, 5000, and 15000 are plotted. With λ increasing, the noise level gets reduced at the expense of a dramatic drop in the peak value of the aluminum

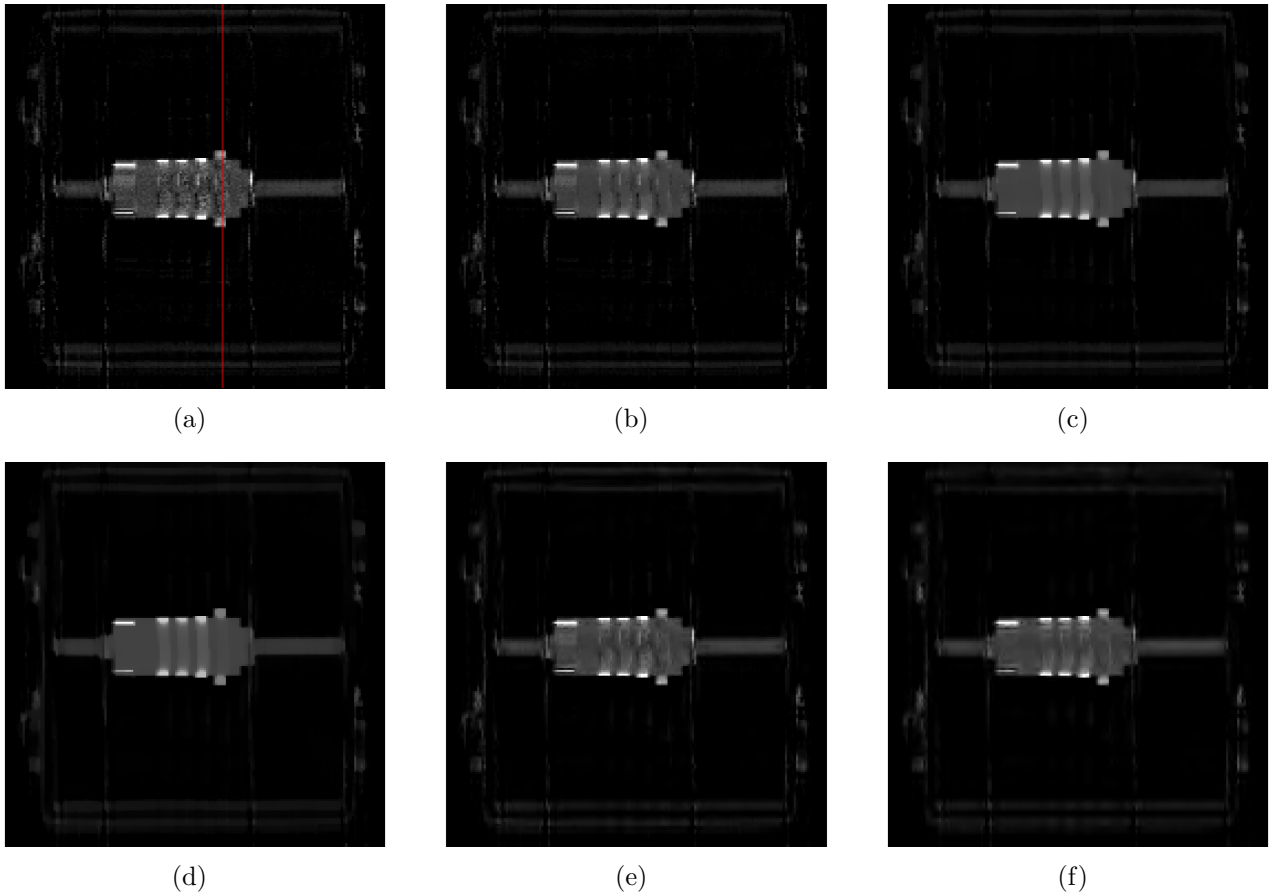


Figure 5.5: Lateral slice no. 186 of the reconstructed volume (a) with unregularized AM, (b) with AM and $\lambda = 1000$, (c) with AM and $\lambda = 5000$, (d) with AM and $\lambda = 15000$, (e) with wav-AM and $\gamma = 1000$, (f) with wav-AM and $\gamma = 3000$. Each transverse image has 172 slices. The images are cropped to the center half containing the case.

wrap. In Figure 5.6(b), a detail-amplified version of the profiles is shown. The aluminum wrap with different λ have not only different values but also different widths. Large penalty weights will generate biased reconstructions and lose the edge resolution.

The profiles for the wav-AM algorithm are shown in Figure 5.7. In Figure 5.7(a), profiles have nearly constant peak values with increasing γ. In Figure 5.7, the edges between air and the aluminum wrap, and the edges between the aluminum wrap and the Delrin cylinder have almost the same sharpness and edge resolution. Even with different choices of γ, the mean

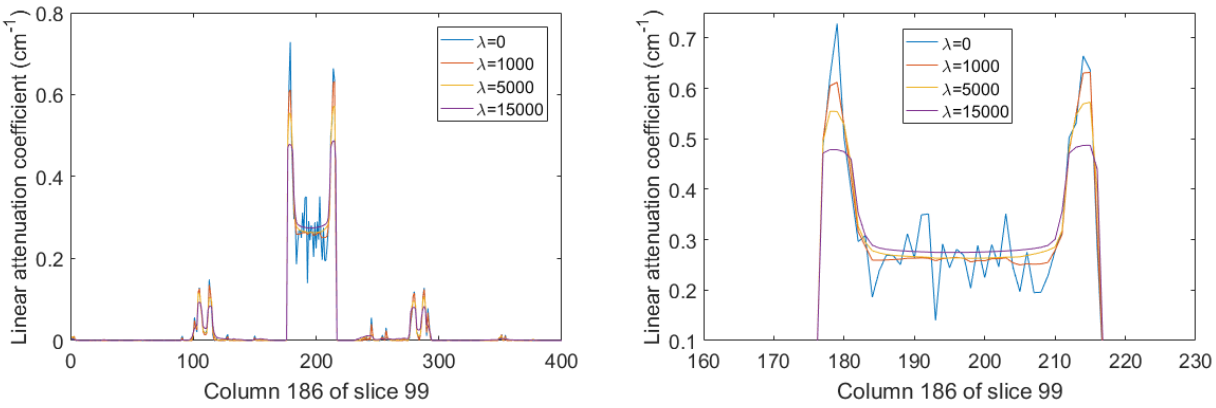


Figure 5.6: Profiles of the reconstructed images from the regularized AM algorithm. (a) profile comparison of column no. 186 of lateral slice no. 99. (b) profile of the center area.

values of the reconstructed image are very close. The bias in the traditional image-domain regularized AM algorithm is reduced with the wav-AM algorithm.

The following figures will show another disadvantage of the AM algorithm compared to wav-AM. In Figure 5.8, the reconstructed images of slice no. 92 are shown. In Figure 5.8(a), with the unregularized AM algorithm, there exist heavy streaks tangent to the boundary of the Delrin cylinder, and dotted noise is visible across the entire region. When the penalty term is added, as shown in Figures 5.8(b), (c), and (d), if the penalty weight is small like $\lambda = 1000$ or 5000, there exist severe artifacts and noise. When λ is increased to 15000, the streak artifacts finally get mostly suppressed. When the wav-AM algorithm is used with $\gamma = 1000$ and $\gamma = 3000$ the streak artifacts are reduced dramatically.

From the bag data experiment, it was shown that the traditional AM algorithm generates biased reconstructions if heavy penalties are used. However, if the penalty weight is not large enough, there still exist unwanted artifacts and heavy noise. To solve this contradictory requirement in the choice of penalty weight, we keep the penalty weight in the traditional

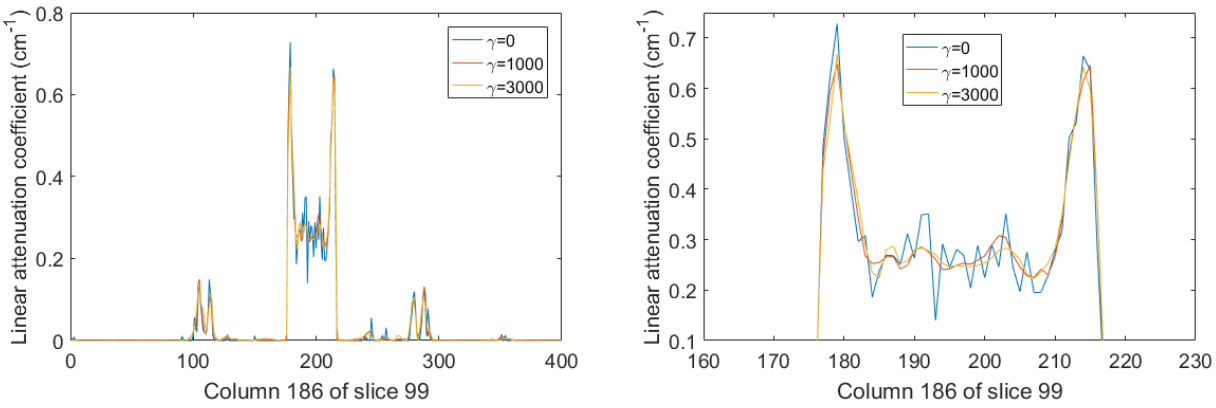


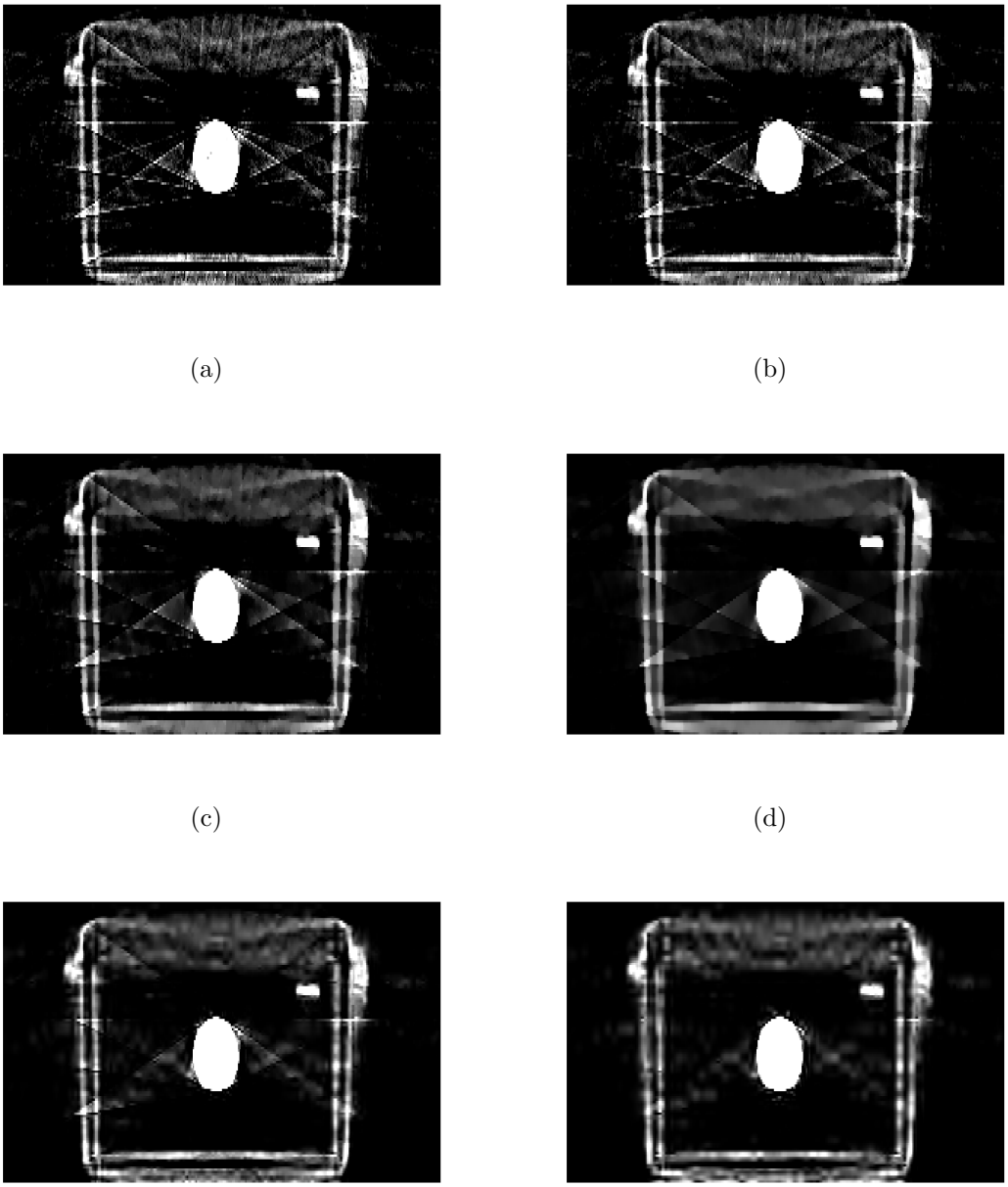
Figure 5.7: Profiles of the reconstructed images from the regularized wav-AM algorithm. (a) profile comparison of column no. 186 of lateral slice no. 99. (b) profile of the center area.

AM algorithm small and add another wavelet penalty term which does not introduce bias to handle the streak artifacts.

5.6 Dual-domain alternating minimization

5.6.1 Dual-domain regularization

As illustrated in [75], although we can use the sparsity of wavelet coefficients as prior knowledge to reduce artifacts and suppress noise, the performance relies heavily on the choice of penalty weight γ. A small weight gives minor improvement in the reconstructed image quality. A large weight introduces blocky artifacts and loses the overall resolution. Finding a proper penalty weight is difficult and time-consuming. There exist no blocky artifacts with the regularized AM algorithm if the edge-preserving penalty is used. However, a small image domain penalty weight cannot produce a smooth reconstruction. Although a large penalty



(a) (b)

(c) (d)

(e) (f)

Figure 5.8: Slice no. 92 of the reconstructed volume (a) with unregularized AM, (b) with AM and $\lambda = 1000$, (c) with AM and $\lambda = 5000$, (d) with AM and $\lambda = 15000$, (e) with wav-AM and $\gamma = 1000$, (f) with wav-AM and $\gamma = 3000$. Each transverse image has 172 slices. The images are cropped to the center half containing the case.

weight can do so, it will generate a biased reconstruction at the same time. A dual domain alternating minimization (dual-AM) algorithm is proposed to solve these contradictory requirements.

The cost function for dual-AM is

$$f(u,\beta) = \sum_i I(d_i||y_i) + \lambda R(u) + \gamma K(\beta), \tag{5.25}$$

where $R(u) = \sum_j \sum_{j\in\mathcal{N}_k} \phi(u_j - u_k)$ is the image-domain neighborhood penalty, and $K(\beta) = \sum_l |\beta_l|$ is the wavelet-domain penalty. Because the wavelet transform matrix Ω and its inverse transform matrix $\hat{\Omega}$ are both orthogonal, this dual-domain optimization problem ca be solved in either the image domain or the wavelet domain by

$$\min_u \sum_i I(d_i||y_i(u)) + \lambda R(u) + \gamma K(\hat{\Omega}u), \tag{5.26}$$

or

$$\min_\beta \sum_i I(d_i||y_i(\beta)) + \lambda R(\Omega u) + \gamma K(\beta) \tag{5.27}$$

These two problems are equivalent and can be interchangeably converted by $\beta = \hat{\Omega}u$ and $u = \Omega\beta$. For the simplicity of computation, problem (5.26) is solved.

Due to the existence of $\hat{\Omega}$, penalty term $K(\hat{\Omega}u)$ is self-coupled. It is hard to use the same decoupling method as in the regularized AM algorithm. Instead, the sparse reconstruction by separable approximation algorithm (SpaRSA) is used [76].

Because $I(d_i||y_i)$ and $R(u)$ are both convex with respect to u, problem (5.26) can be reformulated as

$$\min_u g(u) + \gamma K(\hat{\Omega}u), \tag{5.28}$$

where $g(u) = \sum_i I(d_i||y_i) + \lambda R(u)$. A sequence of iterates $\{u^{(n)} : n = 0\,, 1\,, 2\,, ...\}$ is generated. Problem (5.28) is solved by iteratively solving the following sub-problem,

$$u^{(n+1)} = \arg\min_z (z - u^{(n)}) \frac{\partial g(u)}{\partial u}\Big|_{u=u^{(n)}} + \frac{\alpha_n}{2}||z - u^{(n)}||^2 + \gamma K(\hat{\Omega} z), \tag{5.29}$$

where $\alpha_n > 0$. An equivalent form of the sub-problem (5.29) is

$$u^{(n+1)} = \arg\min_z \frac{1}{2}||z - \hat{u}^{(n+1)}||^2 + \frac{\gamma}{\alpha_n} K(\hat{\Omega} z), \tag{5.30}$$

where

$$\hat{u}^{(n+1)} = u^{(n)} - \frac{1}{\alpha_n} \frac{\partial g(u)}{\partial u}\Big|_{u=u^{(n)}}. \tag{5.31}$$

α_n is taken as the second order derivative of $g(u)$. By this definition,

$$u^{(n+1)} = \arg\min_z \frac{1}{2}||z - \hat{u}^{(n+1)}||^2 + \frac{\gamma}{\alpha_n} K(\hat{\Omega} z), \tag{5.32}$$

where

$$\hat{u}^{(n+1)} = u^{(n)} - \left[\frac{\partial g(u)}{\partial u}\right] / \left[\frac{\partial^2 g(u)}{\partial u^2}\right]\Big|_{u=u^{(n)}} \tag{5.33}$$

is the Newton's update for problem (5.5).

Because Ω is an orthogonal wavelet transform operator, equation (5.32) can be rewritten as

$$u^{(n+1)} = \arg\min_z \frac{1}{2}||\Omega z - \Omega \hat{u}^{(n+1)}||^2 + \frac{\gamma}{\alpha_n} |\hat{\Omega} z|, \tag{5.34}$$

or

$$\beta^{(n+1)} = \arg\min_{\tilde{z}} \frac{1}{2}||\tilde{z} - \hat{\beta}^{(n+1)}||^2 + \frac{\gamma}{\alpha_n} |\tilde{z}|, \tag{5.35}$$

where $\tilde{z}$ and $\hat{\beta}$ represent the wavelet coefficients of z and $\hat{u}$, respectively. It has been demonstrated that the iterative shrinkage thresholding algorithm (ISTA) [77] provides a solution

$$\beta^{(n+1)} = \Big(|\hat{\beta}^{(n+1)} - \frac{\gamma}{\alpha_n}|\Big)_+ \operatorname{sgn}(\hat{\beta}^{(n+1)}) \tag{5.36}$$

In this way problem (5.26) is solved in a two-step scheme. First, a traditional image domain regularized AM update is derived, and then the corresponding wavelet coefficients are updated with (5.36).

Another advantage of the dual-domain AM algorithm is that there is no need for the matrix form of the discrete wavelet transform operator Ω, and there is no need to compute the forward matrix Ψ. A fast wavelet transform can be applied to accelerate the computation of $\beta = \Omega u$.

The dual-domain alternating minimization algorithm is summarized as

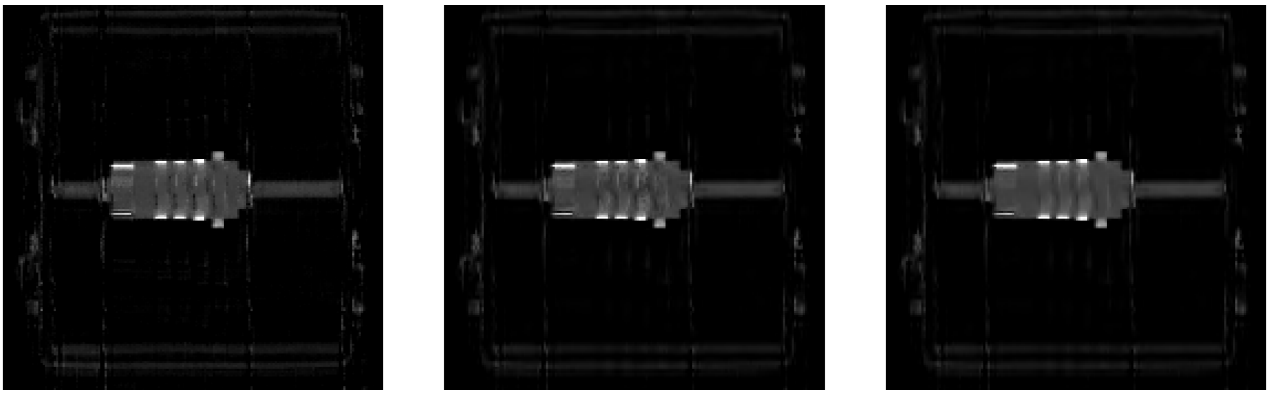
Algorithm 6: Dual-domain alternating minimization algorithm

Initialize $u^{(0)}$.
for $n = 0$ to $N - 1$ **do**
  Solve the image-domain regularized AM with Newton's method to update $\hat{u}^{(n+1)}$.
  Update the wavelet coefficients $\beta^{(n+1)}$ with (5.36).
  Update $u^{(n+1)}$ with $u^{(n+1)} = \Omega\beta^{(n+1)}$.
end

5.6.2 Bag data experiment

In this section, the same bag data as in Section 5.5.4 was used to demonstrate the performance of the dual-AM algorithm. The reconstructed images of lateral slice No. 186 are presented in Figure 5.9. In Figure 5.9(a), the regularized AM algorithm result with $\lambda = 1000$

is shown. In Figure 5.9(b), the regularized wav-AM result with $\gamma = 1000$ is shown. In Figure 5.9(c), reconstructed with the dual-AM algorithm and $\lambda = 1000$, $\gamma = 1000$, the image has the lowest noise level compared with (a) and (b). Compared with the images in Figure 5.5(c) and (d), in which heavy penalty weights $\lambda = 5000$ and $\lambda = 15000$ were used, the image in Figure 5.9 does not suffer the oversmoothing effect.

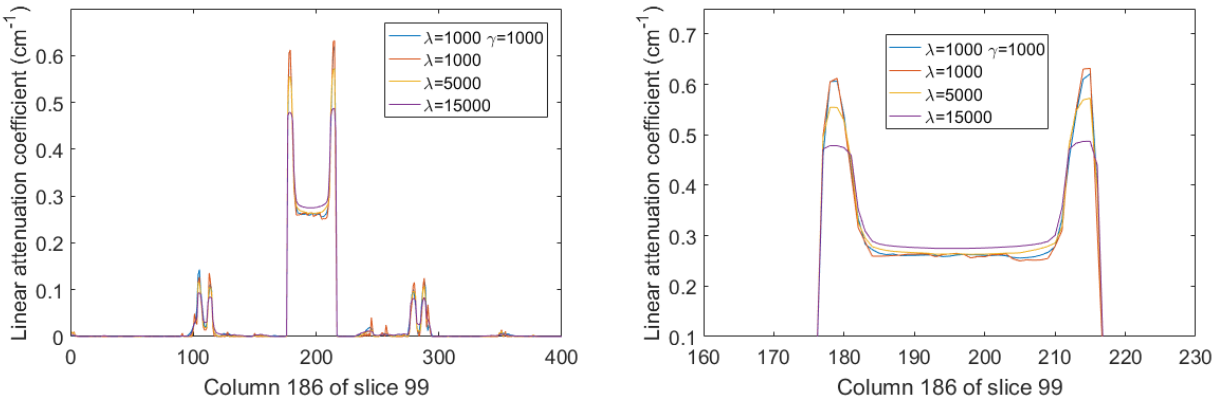


(a) (b) (c)

Figure 5.9: Slice no. 186 of the reconstructed volume: (a) the regularized AM algorithm with $\lambda = 1000$, (b) the regularized wav-AM algorithm with $\gamma = 1000$, (c) the dual-AM algorithm with $\lambda = 1000$ and $\gamma = 1000$.

In Figure 5.10, the profiles of the dual-AM algorithm results are compared with the regularized AM algorithm. In Figure 5.10(a), with $\lambda = 1000$ and $\gamma = 1000$ in the dual-AM algorithm, the peak value of the profile is in agreement with the result of the regularized AM algorithm with $\lambda = 1000$, and is significantly higher than the results with larger penalty weights. In Figure 5.10(b), the center area of the profiles are magnified. The dual-AM algorithm retains good edge preservation on the boundaries between different materials.

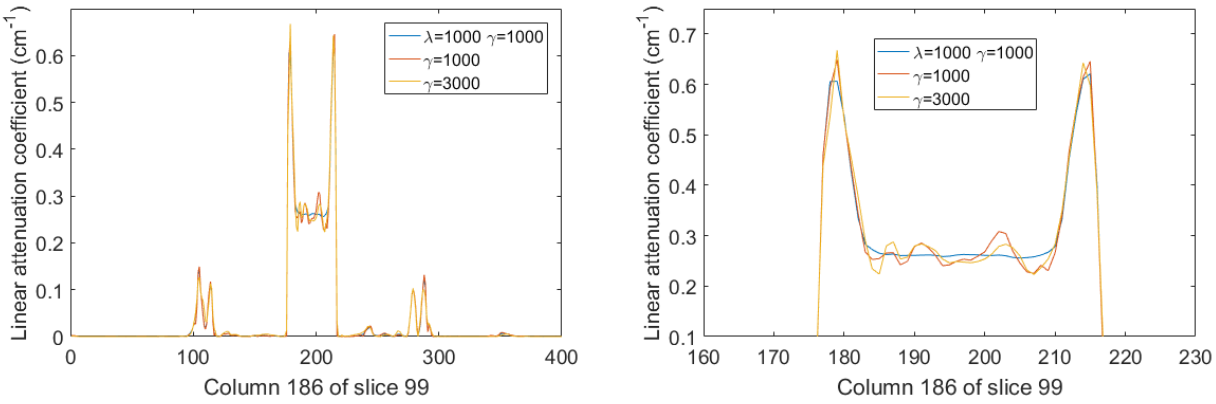
In Figure 5.11, the profiles of reconstructed image column no. 99 are compared with the results of the wav-AM algorithm. In Figure 5.11(a), with $\lambda = 1000$ and $\gamma = 1000$ in the dual-AM algorithm, the peak value is very close to the peak values from the wav-AM algorithm



(a) (b)

Figure 5.10: Profiles of the reconstructed images with the dual-AM algorithm and the regularized AM algorithm: (a) profile comparison of column no. 186 of lateral slice no. 99, (b) profile of the center area.

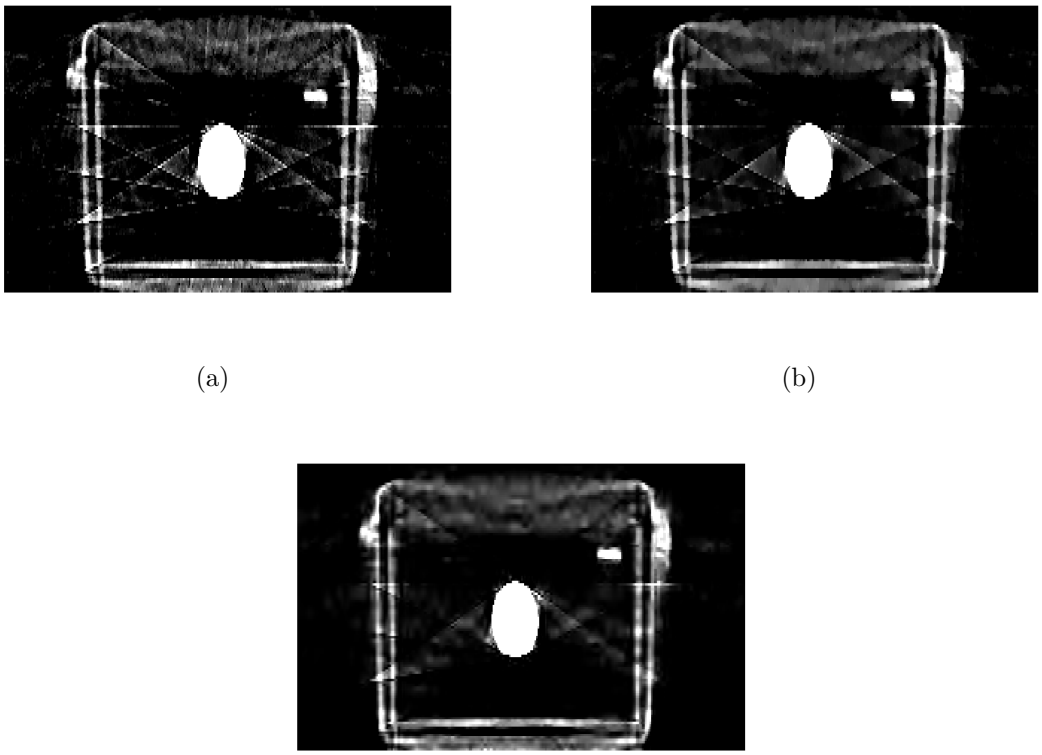
with $\gamma = 1000$ and $\gamma = 3000$. In Figure 5.11(b), in the center area of the profiles, the dual-AM algorithm outperforms the wav-AM algorithm in noise reduction.



(a) (b)

Figure 5.11: Profiles of the reconstructed images from the dual-AM algorithm and the regularized wav-AM algorithm: (a) profile comparison of column no. 186 of lateral slice no. 99, (b) profile of the center area.

The dual-AM algorithm can effectively suppress the artifacts shown in 5.8. In Figures 5.12(a) and (b), it is shown that increasing λ from 1000 to 5000 does not eliminate the streak artifacts. From the experiment in the previous section, $\lambda = 5000$ generates over-smoothed images. In Figure 5.12(c), the streak artifacts are dramatically reduced with the dual-AM algorithm.



(a) (b)

(c)

Figure 5.12: Slice no. 99 of the reconstructed volume: (a) the regularized AM algorithm with $\lambda = 1000$, (b) the regularized algorithm with $\lambda = 5000$, (c) the dual-AM algorithm with $\lambda = 1000$ and $\gamma = 1000$.

From the experiments above, it was shown that the challenge of the contradictory requirements of penalty weight choices is solved by introducing another penalty term in the wavelet

domain. The combination of image-domain penalty and wavelet-domain penalty can reduce the noise and streak artifacts at the same time. With this combination of dual-domain penalties, the penalty weight can be set relatively small, and in this way, the biased estimation and over-smoothing when using the large penalty weight is avoided.

5.6.3 Computation cost comparison

As stated in [48], the determining factor of computation time in the AM algorithm and the wav-AM algorithm is the time used to compute forward and backward projections. The time used to compute these two projections is relevant to the complexity of the forward projector H and Ψ. Usually, H is much sparser than Ψ which leads to the fact that for every iteration, the AM algorithm takes less time than the wav-AM algorithm. This is another reason why in the dual-AM algorithm, the cost function (5.26) is minimized in the image domain instead of the wavelet domain. Still, in the dual-AM algorithm, the dominant time-consuming computation is the image-domain forward and backward projection.

5.7 Summary

A new wavelet domain iterative algorithm was proposed. With the wavelet-domain penalty, wav-AM can generate unbiased reconstruction and reduce noise and artifacts. To incorporate the advantages of the image-domain penalty, the dual-domain alternating minimization algorithm was proposed. From the bag data experiment, the dual-AM algorithm outperforms the single-domain regularized algorithms in noise reduction and artifact suppression.

Acknowledgment

We thank Carl Bosch, Nawfel Tricha, Sean Corrigan, Michael Hollenbeck, and Assaf Mesika of SureScan™ Corporation for their contribution to the collection, formatting, and sharing of the data.

Chapter 6

Penalty weight optimization – Laplace prior variational automatic relevance determination (Lap-VARD)

6.1 Highlights

In this chapter, an algorithm to find the optimal penalty weight choice is proposed. The penalty weight is taken as a parameter of a prior Laplace distribution. Optimization of this parameter using an automatic relevance determination framework results in a balance between the sparsity and accuracy of signal reconstruction. The performance of Lap-VARD is shown through a phantom simulation of x-ray CT.

6.2 Motivation

In the previous chapter, the wavelet-domain penalty was proposed. The wavelet-domain penalty can suppress image noise and reduce artifacts. However, the performance of the wavelet-domain penalty relies heavily on the penalty weight choice. The image-domain penalty weight, which is a direct trade-off between image resolution and noise, has been studied for decades. In the wavelet domain, the relationship between the wavelet-domain penalty weight and image quality is not straightforward. Instead of using an empirical choice of the wavelet-domain penalty weight, Lap-VARD is proposed to find the optimal penalty weight.

6.3 Notation

In this section, the notations used in this chapter are summarized.

- d: measured transmission data.
- y: estimated mean.
- u: reconstructed image.
- β: wavelet transform of the reconstructed image
- I: air-scan source intensity.
- H: system matrix.
- Ω: inverse wavelet transform matrix.

- $\hat{\Omega}$: wavelet transform matrix.
- i: source-detector pair index.
- j: image pixel (voxel) index.
- l: wavelet coefficient index.
- μ: mean of the posterior Laplace distribution
- b: standard deviation of the posterior Laplace distribution

6.4 Lap-VARD

6.4.1 Wavelet-domain penalty

In single energy x-ray CT image reconstruction, under the Poisson noise assumption, the maximum log-likelihood estimation cost function is

$$f(u) = \sum_i -d_i \log q_i(u) + q_i(u), \tag{6.1}$$

where

$$q_i(u) = I_i e^{-\sum_j h_{ij} u_j}, \tag{6.2}$$

and u is the image to reconstruct, d_i is the measured transmission data at source-detector pair i, q_i is the estimated mean of the transmission data at source-detector pair i, I_i is the air-scan intensity measured at source-detector pair i, and h_{ij} is the system matrix element that represents the contribution of pixel j to source-detector pair i.

The wavelet-domain penalty was introduced in Chapter 5. The wavelet sparsity penalty is of the form

$$\gamma \sum_l |\beta_l|, \tag{6.3}$$

where β is the wavelet coefficient of the reconstructed image u. β and u are connected by the transform pair

$$u = \Omega \times \beta, \tag{6.4}$$

and

$$\beta = \hat{\Omega} \times u. \tag{6.5}$$

With this transform pair, the regularized cost function becomes

$$f(\beta) = \sum_i \big[- d_i \log q_i(\beta) + q_i(\beta) \big] + \gamma \sum_l |\beta_l|, \tag{6.6}$$

where

$$q_i(\beta) = I_i e^{-\sum_j h_{ij} \sum_l \omega_{jl} \beta_l} = I_i e^{-\sum_l \psi_{il} \beta_l}. \tag{6.7}$$

The new system matrix $\Psi = H \times \Omega$. Because the values of wavelet coefficients β_l have a large dynamic range, using a single penalty weight will not achieve the best performance. Instead, the penalty term is redesigned as

$$\sum_l \gamma_l |\beta_l|. \tag{6.8}$$

The corresponding cost function is

$$f(\beta) = \sum_i \big[- d_i \log q_i(\beta) + q_i(\beta) \big] + \sum_l \gamma_l |\beta_l|. \tag{6.9}$$

Due to the fact that the penalty weight vector γ is in a high dimensional space, there is no empirical way to set the penalty weights. The automatic relevance determination (ARD) [78, 79] framework is used to find the optimal γ.

6.4.2 Proposed method

In sparsity-driven problems, the basic idea is that when the prior knowledge that the image of interest is sparse or sparse in some basis is known, fewer observations are needed to reconstruct the image by using the significant components in the image [80]. A universal method to seek a best sparse approximation is by using an L-1 penalty

$$\min_x L(x) = f(x) + \lambda|x|, \tag{6.10}$$

where $L(x)$ is the cost function, and $f(x)$ is the data-fitting term. This is equivalent to the L-1 regularization problem [81]

$$\min_x |x|, \text{ s.t. } f(x) \leq \epsilon. \tag{6.11}$$

In problem (6.10), the optimal solution shrinks to 0 as the penalty weight λ goes to $+\infty$, which means that the sparsity of vector x is controlled by the choice of λ. To get a balance between sparsity and data fitting, the Lap-VARD is proposed to automatically update λ iteratively.

Lap-VARD is inspired by the fact that if the data fitting term in (6.10) is the negative log-likelihood

$$f(x) = -\log p(y|x), \tag{6.12}$$

and let the prior of x be a zero-mean Laplace distribution $p(x;\lambda) = \frac{1}{2\lambda}e^{-\frac{|x|}{\lambda}}$, with $\mathrm{E}[x] = 0$ and $\mathrm{Var}[x] = 2\lambda^2$. Problem (6.10) can be rewritten as

$$L(x) = -\log p(y|x) - \log p(x;\lambda). \tag{6.13}$$

This problem becomes a maximum a posteriori (MAP) estimation [82]. Instead of maximizing the posterior probability over x, λ is taken as a hyper-parameter and the marginal likelihood $p(y;\lambda)$ is maximized,

$$\max_{\lambda} L(\lambda) = \log p(y;\lambda) = \log \int p(y|x)p(x;\lambda)dx. \tag{6.14}$$

Based on (6.9), $p(d|\beta)$ and $p(\beta;\gamma)$ can be defined as

$$\log p(d|\beta) = \sum_i d_i \log q_i(\beta) - q_i(\beta), \tag{6.15}$$

and

$$\log p(\beta;\gamma) = \sum_l \gamma_l |\beta_l|. \tag{6.16}$$

The optimal γ, which is defined as γ^*, is given by

$$\gamma^* = \arg\max_{\gamma} \log p(d;\gamma) = \arg\max_{\gamma} \log \int p(d|\beta)p(\beta;\gamma)d\beta. \tag{6.17}$$

To solve the maximization problem given above, the marginal log-likelihood is rewritten as

$$\log p(d;\gamma) = -\mathrm{E}_{q(\beta)} \log[q(\beta))/p(d,\beta;\gamma)] + \mathrm{D}_{\mathrm{KL}}[q(\beta||p(\beta|d;\gamma)], \tag{6.18}$$

where $\mathrm{E}_{q(\beta)}$ stands for the expected value with respect to $q(\beta)$, and D_{KL} is the Kullback-Leibler (KL) divergence.

A variational method is used to maximize (6.18) iteratively [83]. Because the change of $q(\beta)$ does not affect the value of $\log(d|\gamma)$, at iteration n, we can set $q(\beta)^{(n+1)} = p(\beta|d, \gamma^{(n)})$ such that $\mathrm{D}_{\mathrm{KL}}[q(\beta||p(\beta|d;\gamma)] = 0$. Only the first term in (6.18) , which is called *free variational energy* (FVE), needs to be maximized with respect to γ. The EM algorithm [83] can be viewed as minimizing the FVE function by alternately updating $q(\beta)$ and γ. However, the posterior $p(\beta|d,\gamma)$ has no closed-form expression. As in VARD [79], the form of the posterior distribution is restricted. Here, we set the posterior to be a Laplace distribution $q(\beta) \sim \mathit{Laplace}(\mu, b)$ with $\mathrm{E}[\beta] = \mu$ and $\mathrm{Var}[\beta] = 2b^2$.

The objective function is rewritten as

$$\begin{aligned}
\log p(d|\gamma,\mu,b) &= -\,\mathrm{E}_{q(\beta)} \log[q(\beta)/p(d,\beta;\gamma)] + \mathrm{D}_{\mathrm{KL}}[q(\beta)||p(\beta|d;\gamma)] \\
&= \int q(\beta)\log p(d|\beta)d\beta + \int q(\beta)\log q(\beta)d\beta - \int q(\beta)\log p(\beta,\gamma)d\beta \\
&= \sum_i \Big(\prod_l \frac{1}{1-(b_l\Psi_{il})}\Big) I_i e^{-\sum_l \Psi_{il}\mu_l} + \sum_i d_i\Big(\sum_l \psi_{il}\beta_l\Big) \\
&\quad + \sum_j \frac{1}{\gamma_j}\Big(b_j e^{-\frac{|\mu_j|}{b_j}} + |\mu_j|\Big) + \sum_l \log(2\gamma_l) - \sum_l \log(2b_l). \qquad (6.19)
\end{aligned}$$

$\log p(d|\gamma,\mu,b)$ is convex with respect to μ for fixed b and γ, and is also convex with respect to b for fixed μ and γ. Though $\log p(d|\gamma,\mu,b)$ is not convex with respect to γ for fixed μ and b, there exists only one stationary point for γ, and it is easy to show that this stationary is the global minimizer when μ and b are fixed.

An alternating iterative algorithm is proposed to minimize function (6.19).

Minimization with respect to γ

The minimizer γ^* has an explicit expression

$$\gamma_l^* = b_l e^{-\frac{|\mu_l|}{b_l}} + |\mu_l|. \tag{6.20}$$

Minimization with respect to μ

The terms in $\log p(d|\gamma, \mu, b)$ that contain μ are

$$F(\mu) = \sum_i \Big(\prod_l \frac{1}{1-(b_l \Psi_{il})} \Big) I_i e^{-\sum_l \Psi_{il}\mu_l} + \sum_i d_i \Big(\sum_l \psi_{il}\beta_l \Big) + \sum_j \frac{1}{\gamma_j} \Big(b_j e^{-\frac{|\mu_j|}{b_j}} + |\mu_j| \Big). \tag{6.21}$$

The parameters μ_l are coupled, so to update them in parallel at every iteration, the convex decomposition lemma introduced in previous chapters is used again, and the resulting surrogate function is

$$G(\mu) = \sum_l \Big[\Big(\sum_i \Psi_{il} d_i \mu_l \Big) + \theta_l^+ e^{-Z_0(\mu_l - \hat{\mu}_l)} - \theta_l^- e^{Z_0(\mu_l - \hat{\mu}_l)} + \frac{b_l}{\gamma_l} e^{-\frac{|\mu_l|}{b_l}} + \frac{|\mu_l|}{\gamma_l} \Big], \tag{6.22}$$

where

$$\begin{aligned}
\theta_l^+ =& \sum_{i\in C}\Big(\prod_{i\in C}\frac{1}{1-(b_l\Psi_{il})^2}\Big)\alpha_{il}I_i e^{-\sum_l \Psi_{il}\hat{\mu}_l} \\
\theta_l^- =& \sum_{i\in \hat{C}}\Big(\prod_{i\in \hat{C}}\frac{1}{1-(b_l\Psi_{il})^2}\Big)\alpha_{il}I_i e^{-\sum_l \Psi_{il}\hat{\mu}_l} \\
Z_0 =& \frac{|\Psi_{il}|}{\alpha_{il}} = \max_i \sum_l |\Psi_{il}| \\
C =& \{(i,l)|\Psi_{il} > 0\},\ \hat{C} = \{(i,l)|\Psi_{il} \leq 0\}.
\end{aligned}$$

$\hat{\mu}_l$ is the current estimate. Given the convex surrogate function with decoupled μ_l, several methods are available. To simply account for the discontinuous derivative of $|\mu_l|$, the sub-gradient method is chosen [84].

Minimization with respect to b

The terms in $\log p(d|\gamma,\mu,b)$ that contain b are

$$F(b) = \sum_i \Big(\prod_l \frac{1}{1-(b_l\Psi_{il})}\Big) I_i e^{-\sum_l \Psi_{il}\mu_l} + \sum_j \frac{1}{\gamma_j}\Big(b_j e^{-\frac{|\mu_j|}{b_j}} + |\mu_j|\Big) - \sum_l \log(2b_j). \quad (6.23)$$

The minimizer b_j^* does not have a closed-form expression. The parameters b_j are coupled, so the convex decomposition lemma is used again to generated the surrogate function

$$G(b) = \sum_l \frac{1}{\gamma_l}\Big(b_l e^{-\frac{|\mu_l|}{b_l}}\Big) - \sum_l\sum_i r_{il}Q_{il}(\hat{b})\frac{1}{1-(\tilde{b}_l\Psi_{il})^2} I_i e^{-\sum_l \Psi_{il}\mu_l} - \sum_l \log(2b_l), \quad (6.24)$$

where

$$\begin{aligned}\sum_l r_{il} \leq & 1 \\ Q_{il}(\hat{b}) = & \prod_{j \neq l} \frac{1}{1 - (\hat{b}_j \Psi_{ij})^2} \\ \tilde{b}_{il} = & \hat{b}_l + \frac{b_l - \hat{b}_l}{r_{il}}.\end{aligned} \tag{6.25}$$

With this surrogate function, the Newton's method is used find the minimizer b_j^* at every iteration.

μ is the estimated mean of the wavelet coefficients β, and γ corresponds to the penalty weight. If the inverse wavelet transform is used such that the image is reconstructed with β, the optimal penalty weight and its corresponding image are generated at the same time.

6.5 Phantom simulation

In this section, a phantom simulation is used to demonstrate the performance of Lap-VARD. The simulation data is from Section 5.5.3. A FORBILD phantom [74] with image size 256×256 and pixel size 1 mm $\times$ 1 mm was used. The number of source-detector pairs is 21600. The wavelet transform is chosen to be a Haar wavelet with 3 levels. The dose level is set to be 150 mAs which is typical for clinical systems.

The Lap-VARD result was compared with the unregularized AM algorithm, the image-domain regularized AM algorithm, and the wavelet-domain regularized AM algorithm. The reconstructed images are shown in Figure 6.1. The detail structures in the red boxes are magnified in the upper-left corners.

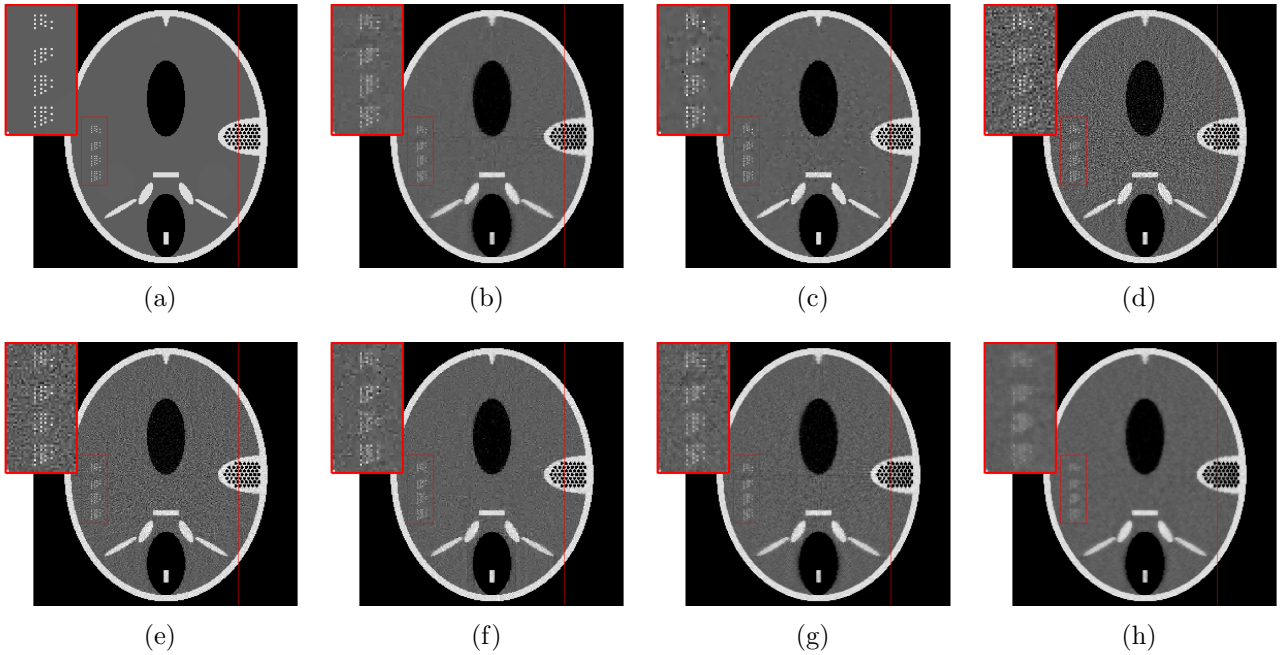
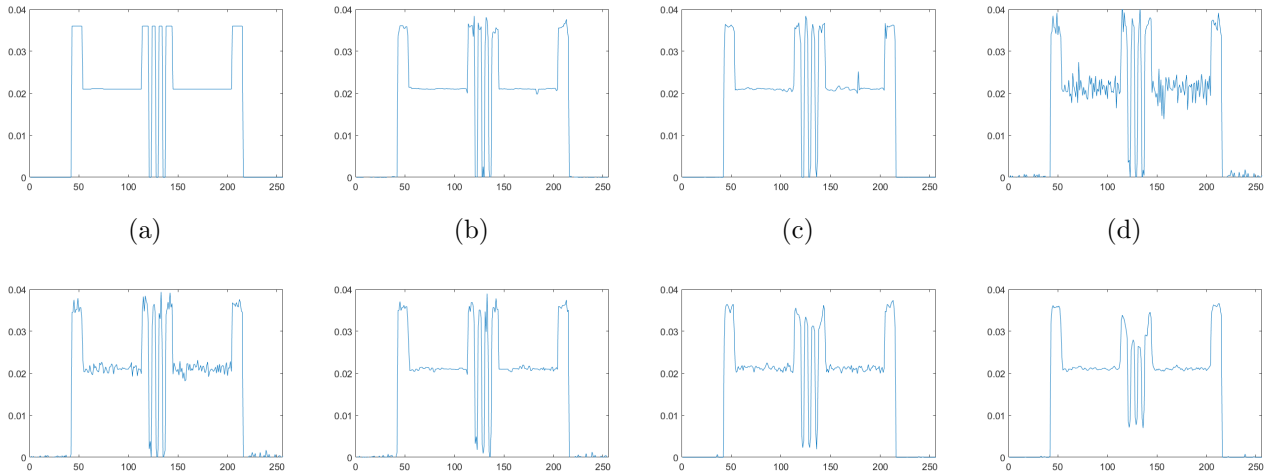


Figure 6.1: Reconstructed images: (a) ground truth, (b) Lap-VARD, (c) VARD, (d) unregularized AM, (e) wavelet-domain regularized AM with penalty weight 300, (f) wavelet-domain regularized AM with penalty weight 1000, (g) image-domain regularized AM with penalty weight 2×10^5, and (h) image-domain regularized AM with penalty weight 6×10^5.

Figure 6.1(a) shows the ground truth for the FORBILD phantom. In Figure 6.1(b), the image was reconstructed with the posterior mean μ by the inverse wavelet transform. The estimated image of the traditional VARD [79] is shown in Figure 6.1(c), which has low noise and good resolution, but there are some isolated bad pixels. Figure 6.1(d) is the unregularized AM result. In Chapter 5, it was shown that because the Haar wavelet transform is orthogonal, the unregularized wavelet AM and the traditional unregularized AM are equivalent, and the result shows a high noise level. Figure 6.1(e) and (f) are wavelet penalized AM algorithm reconstructions with penalty weight 300 and 1000. In (e), the penalty weight is too low such that the noise is still significant. In (f), a proper penalty weight is chosen, and low noise level and high resolution are reached at the same time. Figure 6.1(g) and (h) are image-domain regularized AM results with penalty weights 2×10^5 and 6×10^5. In (g), a proper choice

of the penalty weight generated an image with low noise and high resolution. However, in Figure 6.1(h), when a high penalty weight is used, the image is over-smoothed, and the detail structures are blurred.



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

(e) (f) (g) (h)

Figure 6.2: Profiles of column no. 199: (a) round truth, (b) Lap-VARD, (c) VARD, (d) unregularized AM, (e) wavelet-domain regularized AM with penalty weight 300, (f) wavelet-domain regularized AM with penalty weight 1000, (g) image-domain regularized AM with penalty weight 2×10^5, and (h) image-domain regularized AM with penalty weight 6×10^5.

The profiles of column no. 199 which are highlighted with the red line in Figure 6.1 are shown in Figure 6.2. Figure 6.2(a) shows the ground truth of the profile. In Figure 6.2(b), the reconstruction with Lap-VARD is shown, which gives a quantitatively accurate reconstruction in both the peak-valley contrast and smoothness. In Figure 6.2(c), the profile from traditional VARD is plotted, which has a similar result as the Lap-VARD in smoothness restoration and peak-valley contrast. However, a bad pixel can be seen at the same time. The profile from the unregularized AM algorithm is shown in Figure 6.2(d); the peak-valley contrast is kept, but the noise level is dramatically increased. Figure 6.2(e) plots the result from wavelet regularized AM with penalty weight 300. Still, it shows a high noise level which means the penalty weight is too low. In Figure 6.2(f), with a proper choice of penalty

weight of 1000, the wavelet regularized AM shows comparatively higher noise level than the Lap-VARD result. Figure 6.2(g) is the image-domain regularized AM result with penalty weight 2×10^5, and compared with Lap-VARD result, the noise level is significant. Another severe problem is that the peak and valley values shrink towards their average value. In the over-smoothed result, as shown in Figure 6.2(h), with a large penalty weight 6×10^5, the peak and valley contrast further shrinks and the result is biased.

A quantitative error comparison is summarized in Table 6.1. The root mean square error (RMSE) and peak signal-to-noise ratio (PSNR) with different algorithms are shown. In both metrics, the Lap-VARD algorithm has the best performance.

	Lap-VARD	VARD	AM	wav-AM ($\gamma = 300$)	wav-AM($\gamma = 1000$)	AM($\lambda = 2 \times 10^5$)	AM($\lambda = 6 \times 10^5$)
RMSE	4.86×10^{-4}	7.09×10^{-4}	0.0020	9.45×10^{-4}	7.28×10^{-4}	0.0012	0.0010
PSNR(dB)	38.37	35.42	26.74	32.70	34.93	31.79	29.88

Table 6.1: RMSE and PSNR performances of different algorithms

6.6 Summary

The penalty weight choice has been a hot topic since the idea of using a penalty term was developed. The traditional method to determine the penalty weight heavily relies on an empirical experiment, which is time consuming. In this chapter, a data-driven algorithm is proposed to derive the optimal penalty weight from the ARD framework. At the same time, the image corresponding to the optimal penalty weight choice is reconstructed.

Chapter 7

Conclusions and Future Work

In this dissertation, we have shown that the DEAM algorithm can improve the statistical model in the sinogram-based decomposition method and eliminate the corresponding bias. With a more accurate basis-material image reconstruction, the proposed JSIR framework has the potential to achieve better accuracy and precision on proton stopping power ratio (SPR) estimation, which is important in proton therapy treatment planning, even when the scan dose level is low.

Although it was demonstrated with a simulation that the JSIR framework also generates nearly unbiased estimates in PCCT, the physical model of photon-counting detector is entirely different from the energy-integrating detectors used in dual-energy CT systems. A physical phantom experiment is needed to assess the JSIR framework in the photon-counting CT systems. Another challenge in photon-counting CT systems is that when the radiation dose increases, the pulse pile-up effect dramatically increases and heavily distorts the measurement. Calibration has to be carried out to compensate for the distortion caused by the pile-up effect. The trade-off between the pile-up effect and photon statistics is a crucial consideration in determining the dose level in photon-counting CT systems.

Another dose-related topic in spectral CT is that the PCCT system only requires a single scan, and the DECT system requires high- and low-energy scans. Theoretically, the PCCT system has the potential to cut the radiation dose in half compared with the DECT system. This potential is determined by the photon-counting detector response, which is quite different from the well-studied energy-integrating detectors and is worth further studying.

It was shown that ASAM takes advantage of both the iterative algorithm and the preprocessing method and outperforms them in metal-artifact reduction. However, the metal-artifact reduction performance is parameter related. In clinical settings, the raw data is more complicated than the measured data in the physical phantom experiment. A general guideline for the parameter choice or a data-driven framework to determine the parameters will be the next step to improve the ASAM algorithm.

As was discussed in section 4.7.2, when the scanned object contains heavy metal inserts, the measurements have an extensive dynamic range. In the photon-starved channels, the primary photon statistics are low, and the scatter-to-primary ratio is high. This will result in an underestimation of the attenuation. ASAM has the potential to compensate for the underestimation and give a reference value for the scatter intensity.

A concern about the image-domain penalty is that when the noise level is high, a heavy penalty is required at the cost of loss of resolution and even biased estimation. It was shown that with the wavelet domain penalty, at the same reduction of noise and artifacts, the wav-AM algorithm can generate unbiased reconstructions. Following the wav-AM algorithm, the Lap-VARD algorithm is proposed to determine the optimal wavelet penalty weight which is hard to derive with empirical experiments. In the next step, a physical phantom experiment will be carried out to assess the performance. A challenge that limits the application of the wav-AM and Lap-VARD is the cost of computing forward projection operator Ψ. In section

5.5, Ψ is defined as $\Psi = H \times \Omega$, where H is the CT system matrix and is highly sparse, and Ω is the wavelet transform matrix. The result is that Ψ has huge dimensions and is not sparse. The forward projection requires huge storage space and is time-consuming. Instead of precomputing Ψ and store it locally, an alternative way is to use parallel computing tools like GPUs to compute the forward projection on-the-fly.

Appendix A

Derivation of theorems

A.1 Proof of bias in the sinogram-based decomposition method

In this appendix, we show that the sinogram-based decomposition method generates biased estimates and the bias decreases when the dose level increases.

A.1.1 Mono-energy case

First, we show that the line attenuation coefficient estimation method has a positive bias and this bias decreases when the dose level increases in the mono-energetic case. Assume a mono-energetic x-ray beam with energy E_1 and intensity I_1 passes through a layer of basis material with thickness l. Let $\mu(E_1)$ be the linear attenuation coefficient of basis material 1 at energy E_1. Under the Poisson noise assumption, the measured transmission is given by

$$d_1 \sim Poisson(y_1), \tag{A.1}$$

with mean

$$y_1 = I_1 e^{-\mu(E_1)l}. \tag{A.2}$$

The maximum log-likelihood estimate of the linear attenuation coefficient is given by

$$\hat{\mu}_1 = -\frac{1}{l}\log\frac{d_1}{I_1}. \tag{A.3}$$

The expected value of this MLE value is defined as $\mathrm{E}(\hat{\mu}_1) = \mathrm{E}[-\frac{1}{l}\log\frac{d_1}{I_1}]$. From Jensen's inequality and the fact that $-\log(x)$ is a convex function with respect to x, we have $\mathrm{E}(\hat{\mu_1}) \geq \mu(E_1)$, and the equality is reached when $I_1 \to \infty$.

Let $f(d_1) = -\frac{1}{l}\log\frac{d_1}{I_1}$. Then, the expectation value of $f(d_1)$ is given by

$$\mathrm{E}[f(d_1)] = \mathrm{E}\Big[-\frac{1}{l}\log\frac{d_1}{I_1}\Big]. \tag{A.4}$$

With Taylor series expansion at the point of y_1 to the second order,

$$\mathrm{E}[f(d_1)] \approx f(y_1) + \frac{\partial f(y_1)}{\partial d_1}E[(d_1 - y_1)] + \frac{1}{2}\frac{\partial^2 f(y_1)}{\partial^2 d_1}E[(d_1 - y_1)^2]. \tag{A.5}$$

Due to the fact that y_1 is the mean of Poisson random variable d_1, $E[(d_1 - y_1)] = 0$ and $E[(d_1 - y_1)^2] = y_1$. Then,

$$\begin{aligned}\mathrm{E}\Big[-\log\frac{d_1}{I_1}\Big] \approx& -\frac{1}{l}\log\frac{y_1}{I_1} + \frac{1}{2}\frac{\partial^2 f(y_1)}{\partial^2 d_1}y_1 \\ =&\mu(E_1) + \frac{1}{2}\frac{1}{y_1^2}y_1 \\ =&\mu(E_1) + \frac{1}{2y_1}.\end{aligned} \tag{A.6}$$

The linear attenuation coefficient estimation has a positive bias, $1/2y_1$, which has a inverse relationship with the dose level.

A.1.2 Dual-energy case

Assume two mono-energetic x-ray beams with energies E_1, E_2 and intensities I_1, I_2 pass through the same basis material with thickness l. Let $\mu(E_1)$ and $\mu(E_2)$ be the linear attenuation coefficient of basis material at energies E_1 and E_2. Under the Poisson noise assumption, the statistics of the data are given by

$$d_1 \sim Poisson(y_1),\ d_2 \sim Poisson(y_2), \tag{A.7}$$

with the means

$$y_1 = I_1 e^{-\mu(E_1)l},\ y_2 = I_2 e^{-\mu(E_2)l}. \tag{A.8}$$

The maximum log-likelihood estimate of the thickness l is defined by

$$\hat{l} = \arg\min_l d_1\mu(E_1)l + d_2\mu(E_2)l + I_1 e^{-\mu(E_1)l} + I_2 e^{-\mu(E_2)l}. \tag{A.9}$$

Unlike the mono-energy case, $\hat{l}$ does not have a closed-form solution. However, we can define l_1^* and l_2^* to be the maximum log-likelihood estimates of the mono-energy scans

$$l_1^* = \arg\min_l d_1\mu(E_1)l + I_1 e^{-\mu(E_1)l}, \tag{A.10}$$

and

$$l_2^* = \arg\min_l d_2\mu(E_2)l + I_1 e^{-\mu(E_2)l}, \tag{A.11}$$

both of which have closed-form solutions

$$l_1^* = -\log \frac{d_1}{I_1}/\mu(E_1),\ l_2^* = -\log \frac{d_2}{I_2}/\mu(E_2). \tag{A.12}$$

From the fact that these two mono-energy scans cost functions are both convex with respect to l, we have

$$\min(l_1^*, l_2^*) \leq \hat{l} \leq \max(l_1^*, l_2^*), \tag{A.13}$$

which means that this simplified dual-energy estimate $\hat{l}$ is strictly bounded by the mono-energy estimates l_1^* and l_2^*. From the discussion in the previous mono-energy case, $\hat{l}$ decreases when the dose level increases and converges to the ground truth value l. This proof can be easily extended to continuous spectrum case.

A.1.3 Multi-energy case

The multi-energy case for photon-counting system is more complicated because more than two energy bins measurements are obtained. However, it is also easier to analyze because only one spectrum is used. For each energy on the spectrum, the analysis of section A.1.1 can be applied, and the optimal thickness estimate $l_E^*, E = 1, 2, ..., E_{max}$ can be derived. Using the same analysis in section A.1.2, we have

$$\min(l_E^*, E = 1, 2, ..., E_{max}) \leq \hat{l} \leq \max(l_E^*, E = 1, 2, ..., E_{max}), \tag{A.14}$$

which means that the multi-energy estimate $\hat{l}$ is strictly bounded by the set of mono-energy estimates. We still have the conclusion that $\hat{l}$ decreases when the dose level increases and

converges to the ground truth value l. This proof can be extended to PCCT where the energies across the spectrum are combined into several energy bins.

A.2 Proof of Theorem 2.1

Define a minimization problem

$$\min_p \min_q F = \min_p \min_q \sum_{y,k,E} I(p_k(y,E)||q_k(y,E)) \tag{A.15}$$

s.t.

$$\sum_E p_k(y,E) = d_k(y), \tag{A.16}$$

$$q_k(y,E) = I_{0,k(y)} \sum_E \psi_k(y,E) e^{-\sum_j \mu_j(E) \sum_x h(y|x) c_j(x)}. \tag{A.17}$$

The Lagrange function with respect to constraint (A.16) is

$$\mathcal{L}(p,d) = \sum_{y,k,E} p_k(y,E) \log \frac{p_k(y,E)}{q_k(y,E)} - p_k(y,E) + q_k(y,E) + \lambda \Big(\sum_E p_k(y,E) - d_k(y) \Big) \tag{A.18}$$

Let

$$\frac{\partial \mathcal{L}(p,d)}{\partial p_k(y,E)} = \log p_k(y,E) - \log q_k(y,E) + \lambda = 0, \tag{A.19}$$

and

$$\frac{\partial \mathcal{L}(p,d)}{\partial \lambda} = \sum_E p_k(y,E) - d_k(y) = 0. \tag{A.20}$$

Solving equations (A.19) and (A.20), the results are

$$\lambda = \log \frac{\sum_E q_k(y,E)}{d_k(y)}, \tag{A.21}$$

and

$$p_k(y, E) = q_k(y, E)\frac{d_k(y)}{\sum_E q_k(y, E)}. \tag{A.22}$$

Substitute the minimizer (A.22) back into (A.2),

$$\begin{aligned}
\min_p \min_q F &= \min_p \min_q \sum_{y,k,E} I(p_k(y, E)||q_k(y, E)) \\
&= \min_p \min_q \sum_{y,k,E} p_k(y, E) \log \frac{p_k(y, E)}{q_k(y, E)} - p_k(y, E) + q_k(y, E) \\
&= \min_p \min_q \sum_{y,k,E} q_k(y, E)\frac{d_k(y)}{q_k(y, E)} \log \frac{d_k(y)}{\sum_E q_k(y, E)} - q_k(y, E)\frac{d_k(y)}{\sum_E q_k(y, E)} + q_k(y, E).
\end{aligned} \tag{A.23}$$

If we define

$$Q_k(y) = \sum_E q_k(y, E), \tag{A.24}$$

the above minimization problem becomes

$$\min_p \min_q F = \min_p \min_q \sum_{y,k} d_k(y) \log \frac{d_k(y)}{Q_k(y)} - d_k(y) + Q_k(y). \tag{A.25}$$

Because problem (A.25) is independent of p and $d_k(y)$ is constant, problem (A.25) is equivalent to

$$\min_q F = \min_q \sum_{y,k} Q_k(y) - d_k(y) \log Q_k(y), \tag{A.26}$$

which is problem (2.10).

A.3 Proof of Theorem 3.1

Define a minimization problem

$$\min_p \min_q F = \min_p \min_q \sum_{y,k,E} I(p_k(y,E)||q_k(y,E)) \tag{A.27}$$

s.t.

$$\sum_E p_k(y,E) = d_k(y), \tag{A.28}$$

$$q_k(y,E) = \sum_E I_k(y,E) e^{-\sum_j \mu_j(E) \sum_x h(y|x) c_j(x)}. \tag{A.29}$$

The Lagrange function with respect to constraint (A.28) is

$$\mathcal{L}(p,d) = \sum_{y,k,E} p_k(y,E) \log \frac{p_k(y,E)}{q_k(y,E)} - p_k(y,E) + q_k(y,E) + \lambda\Big(\sum_E p_k(y,E) - d_k(y)\Big). \tag{A.30}$$

Let

$$\frac{\partial \mathcal{L}(p,d)}{\partial p_k(y,E)} = \log p_k(y,E) - \log q_k(y,E) + \lambda = 0 \tag{A.31}$$

and

$$\frac{\partial \mathcal{L}(p,d)}{\partial \lambda} = \sum_E p_k(y,E) - d_k(y) = 0. \tag{A.32}$$

Solving equations (A.31) and (A.32) gives

$$\lambda = \log \frac{\sum_E q_k(y,E)}{d_k(y)}, \tag{A.33}$$

and

$$p_k(y,E) = q_k(y,E) \frac{d_k(y)}{\sum_E q_k(y,E)}. \tag{A.34}$$

Substitute the minimizer (A.34) back into (A.27),

$$\begin{aligned}\min_p \min_q F &= \min_p \min_q \sum_{y,k,E} I(p_k(y,E)||q_k(y,E)) \\ &= \min_p \min_q \sum_{y,k,E} p_k(y,E) \log \frac{p_k(y,E)}{q_k(y,E)} - p_k(y,E) + q_k(y,E) \\ &= \min_p \min_q \sum_{y,k,E} q_k(y,E) \frac{d_k(y)}{q_k(y,E)} \log \frac{d_k(y)}{\sum_E q_k(y,E)} - q_k(y,E) \frac{d_k(y)}{\sum_E q_k(y,E)} + q_k(y,E).\end{aligned} \tag{A.35}$$

If we define

$$Q_k(y) = \sum_E q_k(y,E), \tag{A.36}$$

the above minimization problem becomes

$$\min_p \min_q F = \min_p \min_q \sum_{y,k} d_k(y) \log \frac{d_k(y)}{Q_k(y)} - d_k(y) + Q_k(y). \tag{A.37}$$

Because problem (A.37) is independent of p and $d_k(y)$ is constant, problem (A.37) is equivalent to

$$\min_q F = \min_q \sum_{y,k} Q_k(y) - d_k(y) \log Q_k(y), \tag{A.38}$$

which is problem (3.8).

Bibliography

[1] L. A. Feldkamp, L. Davis, and J. W. Kress, "Practical cone-beam algorithm," *JOSA A*, vol. 1, no. 6, pp. 612–619, 1984.

[2] G. M. Lasio, B. R. Whiting, and J. F. Williamson, "Statistical reconstruction for x-ray computed tomography using energy-integrating detectors," *Physics in medicine & biology*, vol. 52, no. 8, p. 2247, 2007.

[3] J. A. O'Sullivan, D. L. Snyder, and B. R. Whiting, "Alternating minimization algorithms for transmission tomography using energy detectors," in *Proc. Conf. Record of the Thirty-Sixth Asilomar Conf. Signals, Systems and Computers*, vol. 1, Nov. 2002, pp. 144–147.

[4] J. A. O'Sullivan and J. Benac, "Alternating minimization algorithms for transmission tomography," *IEEE Transactions on Medical Imaging*, vol. 26, no. 3, pp. 283–297, 2007.

[5] J. F. Williamson, S. Li, S. Devic, B. R. Whiting, and F. A. Lerma, "On two-parameter models of photon cross sections: Application to dual-energy CT imaging," *Medical physics*, vol. 33, no. 11, pp. 4115–4129, 2006.

[6] G. Vilches-Freixas, V. T. Taasti, L. P. Muren, J. B. B. Petersen, J. M. Létang, D. C. Hansen, and S. Rit, "Comparison of projection- and image-based methods for proton stopping power estimation using dual energy CT," *Physics and Imaging in Radiation Oncology*, vol. 3, pp. 28–36, 2017.

[7] J.-É. Tremblay, S. Bedwani, and H. Bouchard, "A theoretical comparison of tissue parameter extraction methods for dual energy computed tomography," *Medical physics*, vol. 41, no. 8 Part1, 2014.

[8] B. R. Whiting, "Signal statistics in x-ray computed tomography," in *Medical Imaging 2002: Physics of Medical Imaging*, vol. 4682. International Society for Optics and Photonics, 2002, pp. 53–61.

[9] L. Fu, T.-C. Lee, S. M. Kim, A. M. Alessio, P. E. Kinahan, Z. Chang, K. Sauer, M. K. Kalra, and B. De Man, "Comparison between pre-log and post-log statistical models in ultra-low-dose CT reconstruction," *IEEE transactions on medical imaging*, vol. 36, no. 3, pp. 707–720, 2017.

[10] A. R. De Pierro, "A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography," *IEEE Transactions on Medical Imaging*, vol. 14, no. 1, pp. 132–137, Mar. 1995.

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

[12] J. A. Fessler, E. P. Ficaro, N. H. Clinthorne, and K. Lange, "Grouped-coordinate ascent algorithms for penalized-likelihood transmission image reconstruction," *IEEE Transactions on Medical Imaging*, vol. 16, no. 2, pp. 166–175, Apr. 1997.

[13] Y. Wang, J. Yang, W. Yin, and Y. Zhang, "A new alternating minimization algorithm for total variation image reconstruction," *SIAM Journal on Imaging Sciences*, vol. 1, no. 3, pp. 248–272, 2008.

[14] J. M. Boone, "Equivalent spectra as a measure of beam quality," *Medical physics*, vol. 13, no. 6, pp. 861–868, 1986.

[15] J. M. Boone, "The three parameter equivalent spectra as an index of beam quality," *Medical physics*, vol. 15, no. 3, pp. 304–310, 1988.

[16] J. D. Evans, B. R. Whiting, D. G. Politte, J. A. O'Sullivan, P. F. Klahr, and J. F. Williamson, "Experimental implementation of a polyenergetic statistical reconstruction algorithm for a commercial fan-beam CT scanner," *Physica Medica*, vol. 29, no. 5, pp. 500–512, 2013.

[17] R. Birch and M. Marshall, "Computation of bremsstrahlung x-ray spectra and comparison with spectra measured with a Ge (Li) detector," *Physics in Medicine & Biology*, vol. 24, no. 3, p. 505, 1979.

[18] D. Han, J. V. Siebers, and J. F. Williamson, "A linear, separable two-parameter model for dual energy CT imaging of proton stopping power computation," *Medical physics*, vol. 43, no. 1, pp. 600–612, 2016.

[19] S. Zhang, D. Han, D. G. Politte, J. F. Williamson, and J. A. O'Sullivan, "Impact of joint statistical dual-energy CT reconstruction of proton stopping power images: Comparison to image-and sinogram-domain material decomposition approaches," *Medical physics*, vol. 45, no. 5, pp. 2129–2142, 2018.

[20] "American national standard for evaluating the image quality of x-ray computed tomography (CT) security-screening systems," *ANSI N42.45-2011*, pp. 1–58, May 2011.

[21] K. Mechlem, S. Ehn, T. Sellerer, E. Braig, D. Münzel, F. Pfeiffer, and P. B. Noël, "Joint statistical iterative material image reconstruction for spectral computed tomography using a semi-empirical forward model," *IEEE transactions on medical imaging*, vol. 37, no. 1, pp. 68–80, 2018.

[22] T. R. Johnson, "Dual-energy CT: General principles," *American Journal of Roentgenology*, vol. 199, no. 5 supplement, pp. S3–S8, 2012.

[23] Y. Chen, J. A. O'Sullivan, D. G. Politte, J. D. Evans, D. Han, B. R. Whiting, and J. F. Williamson, "Line integral alternating minimization algorithm for dual-energy x-ray CT image reconstruction." *IEEE Transactions on Medical Imaging*, vol. 35, pp. 685–698, Feb 2016.

[24] H. Menzel, C. Clement, and P. DeLuca, "ICRP publication 110. Realistic reference phantoms: An ICRP/ICRU joint effort. A report of adult reference computational phantoms," *Annals of the ICRP*, vol. 39, no. 2, p. 1, 2009.

[25] R. E. Alvarez and A. Macovski, "Energy-selective reconstructions in x-ray computerised tomography," *Physics in Medicine & Biology*, vol. 21, no. 5, p. 733, 1976.

[26] R. Yuan, W. P. Shuman, J. P. Earls, C. J. Hague, H. A. Mumtaz, A. Scott-Moncrieff, J. D. Ellis, J. R. Mayo, and J. A. Leipsic, "Reduced iodine load at CT pulmonary angiography with dual-energy monochromatic imaging: comparison with standard CT pulmonary angiography—a prospective randomized trial," *Radiology*, vol. 262, no. 1, pp. 290–297, 2012.

[27] L. Yu, J. A. Christner, S. Leng, J. Wang, J. G. Fletcher, and C. H. McCollough, "Virtual monochromatic imaging in dual-source dual-energy CT: Radiation dose and image quality," *Medical physics*, vol. 38, no. 12, pp. 6371–6379, 2011.

[28] J. Schlomka, E. Roessl, R. Dorscheid, S. Dill, G. Martens, T. Istel, C. Bäumer, C. Herrmann, R. Steadman, G. Zeitler *et al.*, "Experimental feasibility of multi-energy photon-counting K-edge imaging in pre-clinical computed tomography," *Physics in Medicine & Biology*, vol. 53, no. 15, p. 4031, 2008.

[29] D. P. Clark and C. T. Badea, "Spectral diffusion: An algorithm for robust material decomposition of spectral CT data," *Physics in Medicine & Biology*, vol. 59, no. 21, p. 6445, 2014.

[30] C. Maaß, M. Baer, and M. Kachelrieß, "Image-based dual energy CT using optimized precorrection functions: A practical new approach of material decomposition in image domain," *Medical physics*, vol. 36, no. 8, pp. 3818–3829, 2009.

[31] Y. Zou and M. D. Silver, "Analysis of fast kV-switching in dual energy CT using a pre-reconstruction decomposition technique," in *Medical Imaging 2008: Physics of Medical Imaging*, vol. 6913. International Society for Optics and Photonics, 2008, p. 691313.

[32] T. Flohr, H. Bruder, K. Stierstorfer, M. Petersilka, B. Schmidt, and C. McCollough, "Image reconstruction and image quality evaluation for a dual source CT scanner," *Medical physics*, vol. 35, no. 12, pp. 5882–5897, 2008.

[33] T. Flohr, K. Stierstorfer, S. Ulzheimer, H. Bruder, A. Primak, and C. McCollough, "Image reconstruction and image quality evaluation for a 64-slice CT scanner with-flying focal spot," *Medical physics*, vol. 32, no. 8, pp. 2536–2547, 2005.

[34] S. Feuerlein, E. Roessl, R. Proksa, G. Martens, O. Klass, M. Jeltsch, V. Rasche, H.-J. Brambs, M. H. Hoffmann, and J.-P. Schlomka, "Multienergy photon-counting K-edge imaging: potential for improved luminal depiction in vascular imaging," *Radiology*, vol. 249, no. 3, pp. 1010–1016, 2008.

[35] K. Taguchi, M. Zhang, E. C. Frey, X. Wang, J. S. Iwanczyk, E. Nygard, N. E. Hartsough, B. M. Tsui, and W. C. Barber, "Modeling the performance of a photon counting x-ray detector for CT: Energy response and pulse pileup effects," *Medical physics*, vol. 38, no. 2, pp. 1089–1102, 2011.

[36] D. Pan, E. Roessl, J.-P. Schlomka, S. D. Caruthers, A. Senpan, M. J. Scott, J. S. Allen, H. Zhang, G. Hu, P. J. Gaffney *et al.*, "Computed tomography in color: Nano K-enhanced spectral CT molecular imaging," *Angewandte Chemie International Edition*, vol. 49, no. 50, pp. 9635–9639, 2010.

[37] K. Taguchi, E. C. Frey, X. Wang, J. S. Iwanczyk, and W. C. Barber, "An analytical model of the effects of pulse pileup on the energy spectrum recorded by energy resolved photon counting x-ray detectors," *Medical physics*, vol. 37, no. 8, pp. 3957–3969, 2010.

[38] J. Cammin, S. Srivastava, Q. Tang, W. Barber, J. Iwanczyk, N. Hartsough, and K. Taguchi, "Compensation of nonlinear distortions in photon-counting spectral CT: Deadtime loss, spectral response, and beam hardening effects," in *Medical Imaging 2012: Physics of Medical Imaging*, vol. 8313. International Society for Optics and Photonics, 2012, p. 83131T.

[39] H. Ding and S. Molloi, "Image-based spectral distortion correction for photon-counting x-ray detectors," *Medical physics*, vol. 39, no. 4, pp. 1864–1876, 2012.

[40] D. Vavrik and J. Jakubek, "Material analysis using characteristic transmission spectra," in *Nuclear Science Symposium Conference Record, 2008. NSS'08. IEEE.* IEEE, 2008, pp. 2452–2455.

[41] J. Dickmann, J. Maier, S. Sawall, T. Thüring, S. Gkoumas, C. Brönnimann, and M. Kachelrieb, "A count rate-dependent method for spectral distortion correction in photon counting CT," in *Medical Imaging 2018: Physics of Medical Imaging*, vol. 10573. International Society for Optics and Photonics, 2018, p. 1057311.

[42] J. Cammin, J. Xu, W. C. Barber, J. S. Iwanczyk, N. E. Hartsough, and K. Taguchi, "A cascaded model of spectral distortions due to spectral response effects and pulse pileup effects in a photon-counting x-ray detector for CT," *Medical physics*, vol. 41, no. 4, 2014.

[43] K. Taguchi, K. Stierstorfer, C. Polster, O. Lee, and S. Kappler, "Spatio-energetic cross-talk in photon counting detectors: Numerical detector model (Pc TK) and workflow for CT image quality assessment," *Medical physics*, vol. 45, no. 5, pp. 1985–1998, 2018.

[44] K. Lange, R. Carson *et al.*, "EM reconstruction algorithms for emission and transmission tomography," *J Comput Assist Tomogr*, vol. 8, no. 2, pp. 306–16, 1984.

[45] J. Hsieh *et al.*, "Computed tomography: Principles, design, artifacts, and recent advances," SPIE Bellingham, WA, 2009.

[46] G. N. Hounsfield, "Computed medical imaging," *Medical physics*, vol. 7, no. 4, pp. 283–290, 1980.

[47] G. Zentai, "X-ray imaging for homeland security," *International Journal of Signal and Imaging Systems Engineering*, vol. 3, no. 1, pp. 13–20, 2010.

[48] S. Degirmenci, D. G. Politte, C. Bosch, N. Tricha, and J. A. O'Sullivan, "Acceleration of iterative image reconstruction for x-ray imaging for security applications," in *SPIE/IS&T Electronic Imaging*. International Society for Optics and Photonics, 2015, pp. 94 010C–94 010C.

[49] F. E. Boas and D. Fleischmann, "Evaluation of two iterative techniques for reducing metal artifacts in computed tomography," *Radiology*, vol. 259, pp. 894–902, Jun 2011.

[50] E. Meyer, R. Raupach, M. Lell, B. Schmidt, and M. Kachelrieβ, "Normalized metal artifact reduction (NMAR) in computed tomography," *Med. Phys.*, vol. 37, pp. 5482–5493, Oct 2010.

[51] D. L. Snyder, J. A. O'Sullivan, R. J. Murphy, D. G. Politte, B. R. Whiting, and J. F. Williamson, "Image reconstruction for transmission tomography when projection data are incomplete," *Phys. Med. Biol.*, vol. 51, pp. 5603–5619, Nov 2006.

[52] J. L. Prince and J. M. Links, *Medical imaging signals and systems*. Pearson Prentice Hall, Upper Saddle River, NJ, 2006.

[53] F. E. Boas and D. Fleischmann, "CT artifacts: Causes and reduction techniques," *Imaging in Medicine*, vol. 4, no. 2, pp. 229–240, 2012.

[54] G. H. Glover, "Compton scatter effects in CT reconstructions," *Med. Phys.*, vol. 9, no. 6, pp. 860–867, 1982. [Online]. Available: http://dx.doi.org/10.1118/1.595197

[55] W. A. Kalender, R. Hebel, and J. Ebersberger, "Reduction of CT artifacts caused by metallic implants," *Radiology*, vol. 164, pp. 576–577, Aug 1987.

[56] W. J. H. Veldkamp, R. M. S. Joemai, A. J. van der Molen, and J. Geleijns, "Development and validation of segmentation and interpolation techniques in sinograms for metal artifact suppression in CT," *Med. Phys.*, vol. 37, pp. 620–628, Feb 2010.

[57] D. Prell, Y. Kyriakou, M. Beister, and W. A. Kalender, "A novel forward projection-based metal artifact reduction method for flat-detector computed tomography," *Physics in medicine and biology*, vol. 54, pp. 6575–6591, Nov 2009.

[58] T. Li, X. Li, J. Wang, J. Wen, H. Lu, J. Hsieh, and Z. Liang, "Nonlinear sinogram smoothing for low-dose x-ray CT," *IEEE Transactions on Nuclear Science*, vol. 51, no. 5, pp. 2505–2513, 2004.

[59] Y. Zhang, H. Yan, X. Jia, J. Yang, S. B. Jiang, and X. Mou, "A hybrid metal artifact reduction algorithm for x-ray CT," *Medical physics*, vol. 40, no. 4, 2013.

[60] A. C. Kak and M. Slaney, *Principles of computerized tomographic imaging*. IEEE press New York, 1988.

[61] B. De Man, J. Nuyts, P. Dupont, G. Marchal, and P. Suetens, "Metal streak artifacts in x-ray computed tomography: A simulation study," in *Nuclear Science Symposium, 1998. Conference Record. 1998 IEEE*, vol. 3. IEEE, 1998, pp. 1860–1865.

[62] ——, "Reduction of metal streak artifacts in x-ray computed tomography using a transmission maximum a posteriori algorithm," *IEEE transactions on nuclear science*, vol. 47, no. 3, pp. 977–981, 2000.

[63] J. F. Williamson, B. R. Whiting, J. Benac, R. J. Murphy, G. J. Blaine, J. A. O'Sullivan, D. G. Politte, and D. L. Snyder, "Prospects for quantitative computed tomography imaging in the presence of foreign metal bodies using statistical image reconstruction," *Medical physics*, vol. 29, no. 10, pp. 2404–2418, 2002.

[64] X. Zhang, J. Wang, and L. Xing, "Metal artifact reduction in x-ray computed tomography (CT) by constrained optimization," *Medical physics*, vol. 38, no. 2, pp. 701–711, 2011.

[65] D. L. Snyder, J. A. O'Sullivan, B. R. Whiting, R. J. Murphy, J. Benac, J. A. Cataldo, D. G. Politte, and J. F. Williamson, "Deblurring subject to nonnegativity constraints when known functions are present with application to object-constrained computerized tomography." *IEEE transactions on medical imaging*, vol. 20, pp. 1009–1017, Oct 2001.

[66] H. Erdogan and J. A. Fessler, "Ordered subsets algorithms for transmission tomography." *Physics in Medicine and Biology*, vol. 44, pp. 2835–2851, Nov. 1999.

[67] H. M. Hudson and R. S. Larkin, "Accelerated image reconstruction using ordered subsets of projection data," *IEEE transactions on medical imaging*, vol. 13, no. 4, pp. 601–609, 1994.

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

[69] L. I. Rudin and S. Osher, "Total variation based image restoration with free local constraints," in *Proc. 1st Int. Conf. Image Processing*, vol. 1, Nov. 1994, pp. 31–35 vol.1.

[70] J. A. Fessler, "Statistical image reconstruction methods for transmission tomography," *Handbook of medical imaging*, vol. 2, pp. 1–70, 2000.

[71] S. G. Chang, B. Yu, and M. Vetterli, "Adaptive wavelet thresholding for image denoising and compression," *IEEE Transactions on Image Processing*, vol. 9, no. 9, pp. 1532–1546, Sep. 2000.

[72] Q. Xu, D. Yang, J. Tan, A. Sawatzky, and M. A. Anastasio, "Accelerated fast iterative shrinkage thresholding algorithms for sparsity-regularized cone-beam CT image reconstruction," *Medical Physics*, vol. 43, p. 1849, Apr. 2016.

[73] S. Degirmenci, J. A. O'Sullivan, and D. G. Politte, "Multiresolution approach to acceleration of iterative image reconstruction for x-ray imaging for security applications," in *Proceedings of The 13th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*, 2015.

[74] Z. Yu, F. Noo, F. Dennerlein, A. Wunderlich, G. Lauritsch, and J. Hornegger, "Simulation tools for two-dimensional experiments in x-ray computed tomography using the FORBILD head phantom," *Physics in medicine and biology*, vol. 57, no. 13, p. 237, 2012.

[75] J. Lu, J. A. O'Sullivan, and D. G. Politte, "Wavelet regularized alternating minimization algorithm for low dose x-ray CT," in *Proceedings of The 14th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*, 2017, pp. 726–732.

[76] S. J. Wright, R. D. Nowak, and M. A. T. Figueiredo, "Sparse reconstruction by separable approximation," *IEEE Transactions on Signal Processing*, vol. 57, no. 7, pp. 2479–2493, Jul. 2009.

[77] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.

[78] D. P. Wipf and S. S. Nagarajan, "A new view of automatic relevance determination," in *Advances in neural information processing systems*, 2008, pp. 1625–1632.

[79] Y. Kaganovsky, S. Han, S. Degirmenci, D. G. Politte, D. J. Brady, J. A. O'Sullivan, and L. Carin, "Alternating minimization algorithm with automatic relevance determination

for transmission tomography under Poisson noise," *SIAM Journal on Imaging Sciences*, vol. 8, no. 3, pp. 2087–2132, 2015.

[80] R. G. Baraniuk, "Compressive sensing," *IEEE signal processing magazine*, vol. 24, no. 4, 2007.

[81] E. J. Candes, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, vol. 59, no. 8, pp. 1207–1223, 2006.

[82] T. J. Hebert and R. Leahy, "Statistic-based map image-reconstruction from Poisson data using Gibbs priors," *IEEE Transactions on Signal Processing*, vol. 40, no. 9, pp. 2290–2303, 1992.

[83] R. M. Neal and G. E. Hinton, "A view of the EM algorithm that justifies incremental, sparse, and other variants," in *Learning in graphical models*. Springer, 1998, pp. 355–368.

[84] S. Boyd, L. Xiao, and A. Mutapcic, "Subgradient methods," *Lecture notes of EE392o, Stanford University, Autumn Quarter*, vol. 2004, pp. 2004–2005, 2003.

[85] E. Meyer, F. Bergner, R. Raupach, T. Flohr, and M. Kachelrieß, "Normalized metal artifact reduction (NMAR) in computed tomography," in *Proc. IEEE Nuclear Science Symp. Conf. Record (NSS/MIC)*, Oct. 2009, pp. 3251–3255.

[86] C. Lemmens, D. Faul, and J. Nuyts, "Suppression of metal artifacts in CT using a reconstruction procedure that combines map and projection completion," *IEEE transactions on medical imaging*, vol. 28, no. 2, pp. 250–260, 2009.

[87] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin, "An iterative regularization method for total variation-based image restoration," *Multiscale Modeling & Simulation*, vol. 4, no. 2, pp. 460–489, 2005.

[88] D. G. Politte, J. Lu, J. A. O'Sullivan, E. Johnson, and C. Bosch, "Multi-energy penalized maximum-likelihood reconstruction for x-ray security imaging," in *SPIE Defense+ Security*. International Society for Optics and Photonics, 2017, pp. 101 870D–101 870D.

[89] S. Degirmenci, J. A. O'Sullivan, and D. G. Politte, "Stochastic first-order minimization techniques using Jensen surrogates for x-ray transmission tomography," in *IS&T International Symposium on Electronic Imaging 2016 in the Computational Imaging Conference Proceedings*, San Francisco, California, February 14-18, 2016.

[90] J. C. Ye, K. J. Webb, C. A. Bouman, and R. P. Millane, "Optical diffusion tomography by iterative-coordinate-descent optimization in a Bayesian framework," *JOSA A*, vol. 16, no. 10, pp. 2400–2412, 1999.

[91] J. Browne and A. De Pierro, "A row-action alternative to the EM algorithm for maximizing likelihood in emission tomography," *IEEE transactions on medical imaging*, vol. 15, no. 5, pp. 687–699, 1996.

[92] C. A. Bouman and K. Sauer, "A unified approach to statistical tomography using coordinate descent optimization," *IEEE Transactions on image processing*, vol. 5, no. 3, pp. 480–492, 1996.

[93] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Transactions on Information Theory*, vol. 41, no. 3, pp. 613–627, May 1995.

[94] C. Bouman and K. Sauer, "Fast numerical methods for emission and transmission tomographic reconstruction," in *Proc. 27th Conf. Info. Sci. Sys., Johns Hopkins*, 1993, pp. 611–616.

[95] R. Gutjahr, A. F. Halaweish, Z. Yu, S. Leng, L. Yu, Z. Li, S. M. Jorgensen, E. L. Ritman, S. Kappler, and C. H. McCollough, "Human imaging with photon-counting-based CT at clinical dose levels: Contrast-to-noise ratio and cadaver studies," *Investigative radiology*, vol. 51, no. 7, p. 421, 2016.

[96] M. Berger, J. Hubbell, S. Seltzer, J. Chang, J. Coursey, R. Sukumar, D. Zucker, and K. Olsen, "XCOM: Photon cross section database (version 1.5) National Institute of Standards and Technology," *Gaithersburg, MD.: Monday*, vol. 10, p. 39, 2015.

[97] I. Csiszar *et al.*, "Why least squares and maximum entropy? An axiomatic approach to inference for linear inverse problems," *The annals of statistics*, vol. 19, no. 4, pp. 2032–2066, 1991.

Vita

Jingwei Lu

Degrees

Ph.D., Electrical Engineering, Washington University in St. Louis, Missouri, USA, May 2019
M.S., Electrical Engineering, Washington University in St. Louis, Missouri, USA, May 2014
B.S., Electronic Engineering, Fudan University, Shanghai, China, June 2012

Publications

J. Lu, S. Zhang, D. G. Politte, and J. A. O'Sullivan, "Auxiliary sinogram alternating minimization for x-ray CT metal-artifact reduction", in revision for *IEEE trans on medical imaging*.

J. Lu, S. Zhang, J. F. Williamson, T. Webb, D. G. Politte, B. Whiting, and J. A. O'Sullivan, "Experimental assessment of dual-energy CT estimation bias reduction with joint statistical image reconstruction in sinogram domain-based material decomposition," to be presented at *the 61st AAPM annual meeting*, San Antonio, TX, 2019.

J. Lu, S. Zhang, D. G. Politte, and J. A. O'Sullivan, "Low dose photon counting CT reconstruction bias reduction with multi-energy alternating minimization algorithm" to appear in Proceedings of the 15th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine, 2019.

J. Lu, J. A. O'Sullivan, and D. G. Politte, "Wavelet regularized alternating minimization algorithm for low dose x-ray CT," in Proceedings of The 14th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine, 2017, pp. 726-732.

D. G. Politte, **J. Lu**, J. A. O'Sullivan, E. Johnson, and C. Bosch, "Multi-energy penalized maximum-likelihood reconstruction for x-ray security imaging," in SPIE Defense + Security. International Society for Optics and Photonics, 2017, pp. 101 870D-101 870D.

August 2019