IMPROVING STATISTICAL IMAGE RECONSTRUCTION FOR CARDIAC X-RAY COMPUTED TOMOGRAPHY

by

Jang Hwan Cho

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Electrical Engineering: Systems)
in The University of Michigan
2014

Doctoral Committee:

Professor Jeffrey A. Fessler, Chair
Assistant Professor Laura K. Balzano
Professor Douglas C. Noll
Associate Professor Clayton D. Scott

To my wife, Eun Young, and my son, Adam.

ACKNOWLEDGEMENTS

Last 6 years were the happiest period of my life, without any doubt. I was blessed with so many opportunities in both academic and personal lives. I thank God for orchestrating this amazing journey and introducing so many wonderful people to my life.

First and foremost, I would like to thank my advisor, Jeff Fessler, for being the greatest advisor a graduate student could have. He was always supportive, encouraging, and open to new ideas. His dedication to research and teaching have inspired me, and I have learned a great deal from him. It was my honor and privilege to get to know him and work with him. I really enjoyed working with him because he is not only a great scholar, but also one of the nicest people I ever met. I consider him as my lifetime role model for every aspects of my life. If I can have half as much fun and success in my job as Jeff has, I'll consider myself very lucky. I thank Jeff again for everything he has done for me, and I look forward to keep collaborating with him.

I extend my gratitude to my committee, Professor Douglas C. Noll, Professor Clayton D. Scott, and Professor Laura K. Balzano for their expertise and thoughtful feedback. I was blessed with opportunities to take courses taught by all of committee members, and it helped me a lot to build a solid foundation for this work. I thank them again for their wonderful lectures. I would also like to thank J. B. Thibault, Debashish Pal, and Brian Nett from GE Healthcare. It has been a pleasure to work with them on various important problems for CT, which helped me to understand the modality more thoroughly. I am grateful to all my current and former colleagues for their valuable suggestions. Among my colleagues, I send thanks especially to Sathish Ramani, who is now at GE Global Research Center, for his mentoring.

My parents have always believed in me and understood my passion for knowledge. They provided all the necessary resources to help me achieve my goals. I cannot thank them enough for their unconditional love and support. I would especially thank my grandmother who I respect most. Her love and care got me this far. I would also like to thank my parents-in-law, especially for their spiritual support. My friends in both South Korea and Ann Arbor and conversations with them were great comforts for me.

I thank my wife, Eun Young, for always standing by my side. Without her love and dedicated support, I would not have completed this work. She always believed in me even when I couldn't believe in myself. I thank God for sending my son, Adam, to our lives. His presence alone makes

me the happiest father in the world. Pursuing Ph.D. was a lot like becoming a father for the first time. Trial and error was a basic method for solving problems, and all-nighters were inevitable from time to time. When you think you are getting the hang of it, you will soon find out that there are a lot more to come. However, in spite of all the labors and stresses, you fall in love with your creation no matter how imperfect it is. Since now I know the joy of being a researcher and a father, I feel like the second chapter of my life has just begun.

This work was supported in part by NIH grant R01-HL-098686, GE Healthcare and CPU donation from Intel Corporation. I also want to thank Kwanjeong educational foundation and its founder Chong-hwan Lee, honorary chairman of Samyoung Chemical, for supporting my Ph.D. study. Receiving Kwanjeong scholarship was a great honor, and I understand the meaning and the responsibility of being a recipient of Kwanjeong scholarship. I'll do my best to become a proud asset of my country. Finally, I thank Dr. Kwang-Hyun Cho from Korea Advanced Institute of Science and Technology (KAIST) for his mentoring and support, which helped me receive the scholarship and finishing Ph.D. study.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

LIST OF APPENDICES

Appendix

ABSTRACT

Improving Statistical Image Reconstruction for Cardiac X-ray Computed Tomography
by
Jang Hwan Cho

Chair: Jeffrey A. Fessler

X-ray computed tomography (CT) is one of the most widely used imaging modalities for medical diagnosis. Recent advancements in CT scanner technology have led to increased use of CT in various applications. Unfortunately, these technological advances in CT imaging pose new challenges such as increased X-ray radiation dose and complexity of image reconstruction. Statistical image reconstruction methods use realistic models that incorporate the physics of the measurements and the statistical properties of the measurement noise, and they have potential to provide better image quality and dose reduction compared to the conventional filtered back-projection (FBP) method. However, statistical methods face several challenges that should be addressed before they can replace the FBP method universally. Such challenges include substantial computation time, anisotropic and nonuniform spatial resolution and noise properties, and other artifacts. In this thesis, we develop various methods to overcome these challenges of statistical image reconstruction methods.

Rigorous regularization design methods in Fourier domain were proposed to achieve more isotropic and uniform spatial resolution or noise properties. The design framework is general so that users can control the spatial resolution and the noise characteristics of the estimator. Experimental results show the proposed method can achieve its goal with modest computation cost. In addition, a regularization design method based on the hypothetical geometry concept was introduced to improve resolution or noise uniformity. Proposed designs using the new concept effectively improved the spatial resolution or noise uniformity in the reconstructed image. The hypothetical geometry idea is general enough to be applied to other scan geometries.

We investigated various methods to reduce image artifacts in reconstructed images caused by the short-scan geometry. Statistical weighting modification, based on how much each detector

element affects insufficiently sampled region, was proposed to reduce the artifacts without degrading the temporal resolution within the region-of-interest (ROI). We also proposed a new metric to compare the temporal resolution of the proposed method to that of the short-scan reconstruction. Another approach using an additional regularization term, that exploits information from the prior image, was investigated. Both methods effectively removed short-scan artifacts in the reconstructed image. Experimental results revealed advantages and disadvantages of each approach and their combination.

We accelerated the family of ordered-subsets algorithms by introducing a double surrogate so that faster convergence speed can be achieved. Furthermore, we present a variable splitting based algorithm for motion-compensated image reconstruction (MCIR) problem that provides faster convergence compared to the conjugate gradient (CG) method. Experimental results show that our proposed methods can achieve significant acceleration. The method was also extended to joint estimation of the motion parameters and the reconstructed image. A sinogram-based motion estimation method that does not require any additional measurements other than the short-scan amount of data was introduced to provide decent initial estimates for the joint estimation.

Overall, we proposed various methods that have potential to overcome the major challenges of statistical image reconstruction methods. They were evaluated using simulation and real patient data, and showed promising results. Future work will address more detailed investigation and improvements for these methods. Some of these methods can be combined to generate more complete solutions for CT imaging.

CHAPTER I

Introduction

X-ray computed tomography (CT) is one of the most widely used medical imaging modalities for diagnosis. It provides high-resolution images of anatomical structures of a patient for diagnosing various conditions and frequently replaces investigational surgery for treatment planning and monitoring. Recent advancements in CT scanner technology have led to increased use of CT in various applications. For example, state-of-the-art scanners have more than 64 detector rows and are capable of rotating at 0.35 sec per gantry rotation or faster, which allows them to collect full volumetric datasets in seconds [56]. Fast gantry speed and multi-slice detectors triggered the recent growth in cardiac imaging applications with conventional CT scanners. Unfortunately, these technological advances in CT imaging pose new challenges such as increased X-ray radiation dose and complexity of image reconstruction. Concerns regarding the radiation dose have been growing [53, 70, 81, 125], and a lot of research focuses on reducing the patient dose without degrading the diagnostic quality of reconstructed image [137].

Factors affecting the image quality in CT scans are the X-ray source intensity, which affects patient dose, and the image reconstruction algorithm. Reducing the patient dose can be achieved with reduced X-ray source intensity, but reconstructed images will become much noisier compared to that of regular dose unless a more advanced image reconstruction algorithm is used. From a patient dose and image quality perspective, the major trends of the current research in the field are the following: (i) to develop methods that provide improved image quality, such as finer spatial resolution and better visibility of microscopic features, at current X-ray doses [14, 83, 93, 130, 143], (ii) to develop methods that obtain comparable image quality to current standards, but with significantly lower doses [131, 134, 136, 144]. Statistical image reconstruction (SIR), also known as model-based image (or iterative) reconstruction (MBIR), is a method that has potential to solve this problem. This thesis addresses various aspects of statistical image reconstruction methods and explores ways to overcome the challenges that it is currently facing.

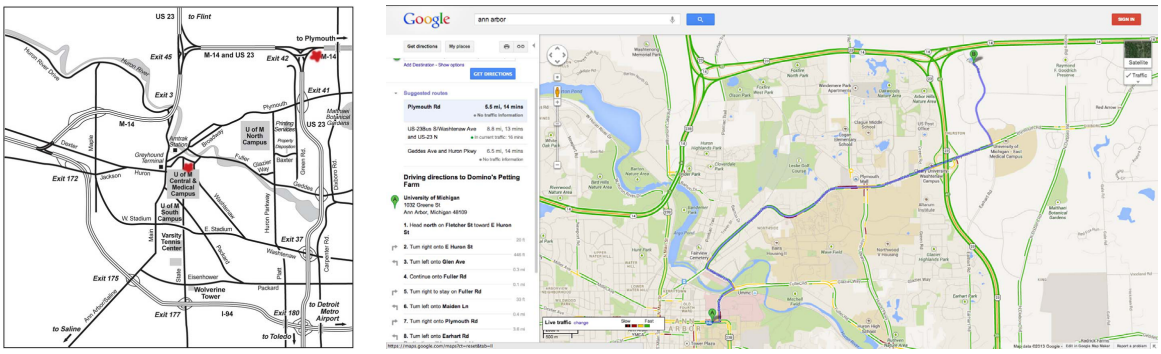
1.1 Statistical image reconstruction in X-ray CT imaging

Image reconstruction is the mathematical process of computing an image of the object being measured in an imaging modality. The process involves the measurements obtained by the modality and their physical characteristics determined by instrument settings. Due to the nature of the data acquisition in a CT scanner, image reconstruction for CT is prone to image artifacts and encounters many challenges to generate a clinically acceptable image. These difficulties are caused by various factors such as the nature of the physics, characteristics of the patient, sub-optimal system design and its usage, and limitations in the current scanning and reconstruction technologies [14, 56]. Many of these problems can be mitigated with advanced hardware having faster gantry speed, better data acquisition system (DAS), etc.. However, modifying the hardware is expensive and not trivial. The role of image reconstruction methods is very critical for this reason.

Today, two major families of image reconstruction methods are being used for X-ray CT: the filtered back-projection (FBP) method and the statistical image reconstruction method. All commercial X-ray CT systems use the FBP method for image reconstruction, and only some of the state-of-the-art scanners support statistical methods (e.g., GE's Veo^{TM}). After being mathematically developed in 1917 [100], FBP method has been the major image reconstruction method in tomography for almost 50 years [30] by extending its use to more advanced scanning geometries such as cone-beam and helical. The model for CT systems used in FBP is very idealized in the sense that it disregards random noise in the measurements. It also assumes the following [56]: (i) the focal spot is infinitely small and can be approximated by a point, (ii) the shape and dimension of each detector cell is ignored, (iii) the size and shape of the reconstruction image voxels are ignored and assumed to be infinitely small. Even though this idealized model does not represent reality, it works well with regular X-ray dose, and FBP produces reasonably good images. However, failure to consider the statistical properties of the measurement noise leads to unacceptable image quality when the dose is significantly lowered. The reconstructed image suffers from severely increased streak artifacts as radiation dose is reduced [62]. Other assumptions may also impact the image quality in other ways.

Statistical image reconstruction methods, on the other hand, use realistic models that incorporate the physics of the measurements and the statistical properties of the measurement noise [40]. Compared to FBP, statistical methods are much more flexible and accurate, because modeling different kinds of physical constraints is easier with these methods. Even though there have been attempts to integrate some aspects of statistical methods into FBP method [141, 142], it is difficult to develop a FBP method with all of the above models.

The difference between the FBP method and the statistical method can be described by using



(a) (b)

Figure 1.1: Maps with (a) limited information[1] and (b) detailed road maps and satellite information[2]

the following example in everyday life. Suppose you arrived at a region where you have never been before and are looking for the hotel that you reserved. If all you have is just a crude map with only major streets and few landmarks (Fig. 1.1 (a)), it is likely that you will have difficulties finding the hotel. Since you have all the necessary information, you can still arrive at the destination, but chances are your search paths would be inefficient and time consuming. On the other hand, if you have a GPS with satellite information (Fig. 1.1 (b)), you'll be able to reach at the hotel in a very efficient manner. The device can provide you with information such as well-modeled road maps, current traffic status, and construction and detour information. Your paths may be comparably efficient if you are in the region with few roads and structures, but they may become very different if you are in a big city. The reconstructed image in CT imaging can be compared to the path that you took in this example. Similar to having a crude map in this example, FBP can certainly provide a reasonable image but its quality may fluctuate depending on various factors like dose. Statistical methods are like having a GPS that can provide various information such that it can produce good images in more flexible and stable manner. Accurate physical models can be compared to the road map data in the GPS, and the current traffic and road status is somewhat analogous to statistical properties of measurements.

Statistical image reconstruction methods have shown their advantages and potential in clinical cases for positron emission tomography (PET). All commercial PET scanners are equipped with statistical image reconstruction methods, and they provide improved image quality over FBP methods [91]. In contrast, even though commercial scanners with statistical methods have been recently launched, the classical FBP image reconstruction method with possible combinations of pre- or post-processing denoising steps [69, 77] is still the dominant reconstruction method in commercial CT scanners. There are several possible reasons for this discrepancy: (i) polyenergetic

[1] ©2008 University of Michigan, MM&D
[2] ©2014 Google

nature of the X-ray source spectrum (ii) much larger data sizes and computational requirements for CT (iii) the FBP image quality in CT is reasonably good compared to that in PET, thus reducing the impetus for using the statistical image reconstruction method (iv) physicians are accustomed to FBP images and their (noisy) texture. However, most of these challenges can be addressed by sophisticated algorithm designs, and as emphasized several times above, the favorable image quality comes at the expense of high X-ray dose. The statistical image reconstruction method has many potential advantages for CT imaging over the conventional FBP method, and thus it is being extensively investigated in various applications.

1.2 Limitations of statistical image reconstruction methods

The feasibility of statistical image reconstruction methods has been demonstrated in [124] in terms of noise, resolution, and artifacts. Other literatures have shown that statistical methods can provide better image quality in these criteria [16, 97, 133]. Such improvements in image quality become more apparent in low-dose scans where FBP reconstruction suffers from increased streak artifacts [62]. However, there are still many factors needing to be addressed to ensure the success of statistical methods in clinical applications. Diagnostic readability of the reconstructed images depends on various characteristics such as texture, resolution, noise, and artifacts. Here, we introduce major challenges that statistical image reconstruction methods are facing currently. We provide possible solutions to each of these problems in the following chapters.

1.2.1 Image artifacts

Due to the complex nature of statistical image reconstruction, some image artifacts appear in the reconstructed image that were not noticeably visible in that of FBP methods. These artifacts can be in the form of shading, correlated texture, etc.. Reasons for these artifacts are not clear, and they may result from combinatorial effects of various factors including regularization design and insufficient sampling. Furthermore, since various kinds of artifacts appear in the reconstructed image such as motion artifacts, beam hardening and streak artifacts [14], it is hard to isolate the cause of each artifact in the image. No matter how local and insignificant an image artifact appears, it can affect the diagnostic readability and impair the integrity of the reconstructed image. In this dissertation, we focus on the short-scan artifact that results from using a standard short-scan scanning protocol for cardiac CT.

1.2.2 Anisotropic and nonuniform spatial resolution and noise characteristics

Spatial resolution or noise properties of reconstructed images are often used as criteria to evaluate imaging systems. Sharp spatial resolution is usually preferred, but it must be balanced with appropriate levels of noise. For diagnostic readability, having uniform and isotropic spatial resolution and noise characteristics are desirable. Unlike many imaging systems that are approximately shift-invariant, CT systems are shift-variant due to their physical nature. For example, in fan-beam CT, more rays are concentrated at the edges of the image compared to the center of the image. In addition, statistical image reconstruction methods use the shift-variant statistics of an object such that the resulting estimators often have shift-variant characteristics.

Spatial resolution and noise properties of reconstructed images can affect the readability of these images, especially for detecting lesions. Anisotropy and nonuniformity of the spatial resolution can distort the shape of certain features in the reconstructed image and also its amplitude [113]. Also, anisotropic and nonuniform noise properties mean different levels of noise for varying voxel locations, and can affect the texture of the reconstructed image. Having these effects in the reconstructed image is obviously undesirable for diagnostic purposes. Therefore, controlling spatial resolution and noise properties for the statistical image reconstruction methods is very important, and has been researched widely [90, 99, 108, 116, 118, 138].

Since uniform spatial resolution and noise characteristics cannot be achieved at the same time, controlling their balance in the reconstructed image is very important. However, there is no clear guideline for such balance because each physician may have different preference for the reconstructed image quality. Therefore, having an ability to control spatial resolution and noise characteristics according to user-selected criteria is highly desirable.

1.2.3 Computational cost

In spite of many potential advantages of the statistical image reconstruction method, the method is not being used ubiquitously in clinical CT imaging yet. One major drawback of the method is the high computational intensity compared to FBP methods. Unlike FBP methods that only require one back-projection and additional post-filtering steps, statistical image reconstruction methods minimize a complex cost function consisting of a data-fitting term incorporating the system model and the measurement statistics, and a regularizer term designed to enforce additional constraints on the reconstructed image such as smoothness and edge-preservation. Minimization of the cost function is generally performed by an iterative algorithm that requires multiple iterations involving forward- and back-projections, and thus computationally extensive. Usually, the computation time taken to execute one iteration of an iterative algorithm is larger than that of a FBP method due to an extra forward-projection operation. In addition, even the highly efficient algorithms require more

than tens of iterations to reach a certain convergence criteria. Reducing the computational cost for statistical methods is thus one of the most active research topics in the field of CT imaging.

1.2.4 Temporal resolution and motion artifacts

Limited temporal resolution and resulting motion artifacts are not caused by using a statistical image reconstruction method, but rather are the result of inherent limitations of the imaging modality and patient motion. Many imaging systems dealing with an object with motion suffer from certain spatial and temporal trade-off in image acquisition, and some of them have difficulties achieving both good spatial and temporal resolutions. Unfortunately, CT falls into this undesirable category. As mentioned earlier, the gantry speed of state-of-the-art CT scanners is 0.35 seconds per rotation or faster. This is enough to obtain a non-cardiac image without having noticeable motion blur as patients can hold their breath during the scan. However, the situation changes for cardiac scans for the following reasons. First, cardiac motion is involuntary. Second, the gantry speed is not fast enough to avoid motion blur especially when the heart rate of the patient is faster than 75 beats-per-minute (BPM). Third, a CT scanner with 64 detector rows, which is still dominant in clinical settings, can cover only part of the heart at one scan, thus requires multiple scans to obtain an image of the entire heart.

There exist some solutions to these issues. Beta blockers can be used for short term reduction of the heart rate prior to cardiac CT scans, and they are safe and effective with well-established protocols. However, beta blockers do not guarantee to lower the heart rate as desired and may have contraindications for certain patients [80]. Some hardware improvements such as faster gantry speed and wider cone angle for cone-beam CT have been introduced by certain manufacturers, and these modifications will be made constantly as technology evolves. However, as previously mentioned, hardware modification, such as advanced gantry or dual source CT, is expensive and nontrivial. Solutions with new scanning protocols or reconstruction methods are thus desired. The statistical image reconstruction method can provide the solution to this problem by incorporating motion models.

1.3 Contributions of this dissertation

This dissertation addresses the challenges of statistical image reconstruction methods for cardiac CT problem.

First of all, we investigated various methods to reduce the image artifacts in reconstructed images obtained from the short-scan approach. We mainly approach the problem from the regularization design point of view. We proposed regularization design methods to obtain isotropic and uniform spatial resolution or noise properties [19–21]. Two different types of regularization

design methods were investigated: one using the Fourier domain approach like previous works in the field, and the other using a new hypothetical geometry concept. Experimental results with both clinical data and phantom study illustrate the advantages and limitations of each method.

Secondly, we approach the short-scan artifact problem by improving the sampling condition at undersampled voxels. Starting from simple extrapolation methods, we investigate methods that use all the available information, such as additional measurements or initial reconstruction, to reduce short-scan artifacts while maintaining the temporal resolution within the region-of-interest (ROI) of the reconstructed image. Statistical weighting modification method and additional prior regularization were proposed and evaluated with phantom simulations for different cone-angles. Furthermore, combination of these methods were investigated.

Finally, we accelerated the statistical image reconstruction method for the cases with and without motion-compensation. We proposed an acceleration scheme for conventional ordered-subsets type methods using a "double surrogate" idea [17] in the case of no motion-compensation. For motion-compensated image reconstruction, we proposed an acceleration scheme using variable splitting approach [24, 25]. In general, motion-compensated image reconstruction relies heavily on the motion estimates, and imperfect estimates leads to residual artifacts. Joint estimation of both motion parameters and reconstructed images have potential to address this issue, but they still require good initial conditions and acceleration scheme. We investigated a sinogram-based motion estimation method [18] that does not require any additional measurements other than the short-scan amount of data. Estimated motion can serve as decent initial estimates for joint estimation of both motion parameters and reconstructed images. Lastly, we extend the variable splitting approach to the joint estimation. Experimental results show that the proposed acceleration algorithms converge faster compared to the conventional ordered-subsets type algorithms and nonlinear conjugate gradient algorithm, respectively.

1.4 Outline of the dissertation

This dissertation is organized as follows.

Chapter II describes models used in the statistical image reconstruction and a mathematical framework for incorporating motion models into the image reconstruction process. Short-scan artifacts are introduced with examples in this chapter.

Chapter III presents regularization designs methods to improve resolution and noise properties and reduce artifacts. Two separate regularization design methods were proposed to achieve more isotropic and uniform spatial resolution and noise properties, respectively.

Chapter IV investigates additional regularization design methods using the hypothetical geometry idea. The designs are more focussed on the uniformity of resolution and noise throughout the

volume, including undersampled regions.

Chapter V introduces methods to reduce short-scan artifacts in the reconstructed image. Statistical weighting modification approach that controls the contribution of the weights associated with the extra measurements was proposed. Additionally, new prior regularization term using structural information from an initial reconstruction was introduced.

Chapter VI presents how to accelerate conventional ordered-subsets type algorithms and motion-compensated image reconstruction methods. A sinogram-based motion estimation algorithm that obtains the motion estimates without any additional measurements other than short-scan data was introduced to obtain feasible initial motion estimates.

Finally, Chapter VII concludes the thesis and outlines the proposed future work.

CHAPTER II

Background

In this chapter, we first review statistical image reconstruction methods for X-ray CT problems. We also describe motion-compensated image reconstruction methods within MBIR framework. Mathematical models and cost functions for these methods will be presented. The models and methods presented in this chapter are general enough to represent different imaging modalities. However, if necessary, more specific assumptions will be made in other chapters. Finally, we introduce short-scan artifacts that are specific to 3D cone-beam CT.

2.1 Statistical image reconstruction (SIR) methods

The purpose in X-ray CT imaging is to accurately estimate the spatial distribution of the linear attenuation coefficient $\mu(\boldsymbol{r}, \mathcal{E})$ of the patient being scanned, where $\boldsymbol{r}$ denotes spatial location and $\mathcal{E}$ parameterizes the dependence on incident photon energy [40]. We need the following steps for any statistical image reconstruction method [37]: (i) object parameterization, (ii) modeling the system physics, (iii) modeling the measurement statistics, (iv) defining a regularization, and (v) designing an iterative algorithm for minimizing the cost function.

2.1.1 Object model

The object can be parameterized using a finite linear combination of basis functions:

$$\mu(\boldsymbol{r}) = \sum_{j=1}^{n_{\mathrm{p}}} x_j B_j(\boldsymbol{r}), \tag{2.1}$$

where n_{p} is the number of coefficients x_j, $\boldsymbol{r}$ denotes the 3D spatial coordinates and B_j denotes the jth basis function. Now, our goal is to reconstruct the coefficients $\boldsymbol{x} = (x_1, \cdots, x_{n_{\mathrm{p}}})$ from the measurements. Notice that we disregarded the dependence of μ on X-ray photon energy $\mathcal{E}$ for simplicity. However, when we are dealing with the patient data, we accounted for the effects of

beam hardening using the standard polynomial water correction [67,85]. Throughout the thesis, we use cubic basis functions corresponding to square voxels, but the methods could be generalized to consider other types of basis functions such as smooth basis functions, so-called "blobs", in [145] at the expense of increased computation. For cone-beam CT, one practical matter to consider is that statistical image reconstruction methods need to account for the entire portion of the object that contributes to the measurements, which means that we also need to reconstruct axial slices that are outside of the detector coverage. This prevents artifacts from propagating into the ROI, and helps reconstructing end slices that are always under-sampled compared to center slices ("long object problem"). Therefore, in practice, n_{p} includes the number of voxels in the diagnostic ROI and also the "padded" slices.

2.1.2 System and measurement models

Let the random variable Y_i represent the signal recorded for the ith ray. We use the following statistical model for the measurements:

$$E[Y_i] = \bar{Y}_i = b_i \exp\left(-\int_{L_i} \mu_0(\boldsymbol{r}) dl\right) + r_i, \tag{2.2}$$

where b_i is the intensity of the ith measurement, r_i denotes the background contributions from factors such as scatter, and crosstalk, L_i is the line between the source and detector for the ith ray, and $\mu_0 \triangleq \mu(\boldsymbol{r}, \mathcal{E}_0)$ is the linear attenuation coefficient at the given source photon energy $\mathcal{E}_0$ [40].

Let $\boldsymbol{y} = (y_1, \cdots, y_{n_{\mathrm{d}}})$ denote the vector of noisy sinogram measurements after taking the logarithm and having been corrected for physical effects such as beam hardening, where n_{d} is the number of measured sinogram values. The mean of each measured sinogram data value y_i has the following linear relation with the attenuation coefficients:

$$E[y_i] = \int a_i(\boldsymbol{r}) \mu(\boldsymbol{r}) d\boldsymbol{r}, \quad i = 1, \cdots, n_{\mathrm{d}}, \tag{2.3}$$

where $a_i(\boldsymbol{r})$ is the system response accouting for effects like finite source and detector size, and other detector characteristics [57]. Substituting (2.1) into (2.3) yields the following linear system

$$E[y_i] = \bar{y}_i = \sum_{j=1}^{n_{\mathrm{p}}} a_{ij} x_j \triangleq [\boldsymbol{A}\boldsymbol{x}]_i, \tag{2.4}$$

where $\boldsymbol{A}$ is the $n_{\mathrm{d}} \times n_{\mathrm{p}}$ system matrix with the following elements

$$a_{ij} \triangleq \int a_i(\boldsymbol{r}) B_j(\boldsymbol{r}) d\boldsymbol{r}. \tag{2.5}$$

In this thesis, we model $\boldsymbol{A}$ using either the separable footprint method [79] or the distance-driven method [32] for forward- and back-projection operators.

2.1.3 Statistical modeling

The measurement noise statistics can be modeled using a probability density function by relating the measurements y_i to their mean values $E[y_i]$. The physical processes in the data acquisition system mainly affect these statistics. For current integrating detectors, the statistics of X-ray measurements is a complicated mixture of compound Poisson photon distribution and Gaussian electronic noise [132]. In practice, the following simple models have been used successfully: (i) a Poisson model for pre-log data Y_i in a discretized form of (2.2) can be written as [77]

$$Y_i \sim \text{Poisson}\{b_i \exp(-[\boldsymbol{A}\boldsymbol{x}]_i) + r_i\}, \tag{2.6}$$

where $i = 1, \cdots, n_{\text{d}}$, and b_i and r_i are assumed to be known. (ii) The quadratic approximation of the negative log-likelihood of (2.6) implies that the post-log data y_i in (2.3) is approximately a Gaussian random variable [38, 107]

$$y_i \sim N\left([\boldsymbol{A}\boldsymbol{x}]_i, \frac{\bar{Y}_i}{(\bar{Y}_i - r_i)^2}\right). \tag{2.7}$$

2.1.4 Cost function formulation for SIR

Statistical image reconstruction methods reconstruct an image $\hat{\boldsymbol{x}}$ from noisy measurements $\boldsymbol{y}$ by minimizing a cost function, $\Psi(\boldsymbol{x})$:

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \Psi(\boldsymbol{x}). \tag{2.8}$$

The cost function consists of a negative log-likelihood or a data-fidelity term $\mathsf{L}(\boldsymbol{x})$, which is obtained from the statistical models described above, and a regularization or a penalty term $\mathsf{R}(\boldsymbol{x})$.

$$\Psi(\boldsymbol{x}) \triangleq \mathsf{L}(\boldsymbol{x}) + \mathsf{R}(\boldsymbol{x}). \tag{2.9}$$

For independent measurements, we can express the log-likelihood term in more convenient form [40]:

$$\mathsf{L}(\boldsymbol{x}) = \sum_{i=1}^{n_{\text{d}}} h_i([\boldsymbol{A}\boldsymbol{x}]_i). \tag{2.10}$$

The function $h_i(l)$ is determined by the selected statistical model. For Poisson model in (2.6), the marginal log-likelihood of the ith measurement is

$$h_i(l) = (b_i \exp(-l) + r_i) - Y_i \log(b_i \exp(-l) + r_i), \tag{2.11}$$

and this function is convex if $r_i = 0$. Based on the Gaussian model in (2.7) and the mean-variance relationship, the negative log-likelihood of the measurements can be approximated by a weighted least-squares (WLS) function:

$$h_i(l) = \frac{1}{2} w_i (l - y_i)^2, \tag{2.12}$$

where the weights w_i should be the reciprocal of the variances of the measurements by the Gauss-Markov theorem:

$$w_i \triangleq \frac{1}{\sigma^2(E[y_i])} = \frac{(\bar{Y}_i - r_i)^2}{\bar{Y}_i}. \tag{2.13}$$

In practice, the mean of the measurements is unknown so typically the weightings are estimated by using a plug-in approach, i.e. $w_i \approx 1/\sigma^2(y_i)$. In this thesis, we mostly focus on using the WLS cost function. One potential drawback of the WLS cost function (2.12) is the use of the logarithm, which may lead to negative measurements for some rays in the very low dose cases. However, this issue is not one of our focuses, so we simply set the weights to zero for such rays assuming that they do not carry much useful information.

A regularization term is added to the log-likelihood term to control the noise and possibly enforce additional requirements to the reconstructed image. Such requirements include edge-perservation, non-negativity, and space-invariant spatial resolution or noise. Furthermore, adding regularization can assist convergence by reducing the condition number of the problem. Absence of regularization may result in undesirably noisy image. We can write a regularization function as the following form:

$$\mathsf{R}(\boldsymbol{x}) = \sum_{j=1}^{n_{\mathrm{p}}} \sum_{l=1}^{N_{\mathrm{l}}} \beta_j^l \psi_l \left((c_l * * * x) [n, m, z] \right), \tag{2.14}$$

where N_{l} is the size of the neighborhood, which is usually 13 in 3D, β_j^l is the regularization parameter that balances between the data-fitting term and the regularizer, ψ_l is the potential function, $***$ denotes 3-D convolution, we define a first-order differencing function that penalizes lth neighbor as

$$c_l[n, m, z] = \frac{\delta[n, m, z] - \delta[n - n_l, m - m_l, z - z_l]}{\sqrt{n_l^2 + m_l^2 + z_l^2}}, \tag{2.15}$$

$\delta[n, m, z]$ is a Kronecker impulse at the location $[0, 0, 0]$, and n_l, m_l, z_l denote the offset of the lth

neighbor. The regularization parameter β_j^l is commonly defined as the following separable form

$$\beta_j^l \triangleq \beta r^l \kappa_j \kappa_{j-j_l}, \tag{2.16}$$

where β is a constant that determines that overall strength of the regularizer, j_l denotes the offset of the lth neighbor in lexicographical order, $\{r^l\}$ are the directional coefficients, and κ_j is the user-defined weight for controlling spatial resolution and noise in the reconstructed image [42]. A typical choice of the directional weights is $r^l = 1, \forall l$, but they also can be selected more sophisticatedly as in [138]. Substituting (2.16) to (2.14) leads to the following expression for a regularizer

$$\mathsf{R}(\boldsymbol{x}) = \beta \sum_{j=1}^{n_{\mathrm{p}}} \kappa_j \sum_{l=1}^{N_{\mathrm{l}}} \kappa_{j-j_l} r^l \psi_l \left((c_l * * * x)\,[n, m, z] \right), \tag{2.17}$$

where, in this case, β is usually referred to as the regularization parameter.

Selection of the potential function ψ_l influences the quality and the characteristics of the reconstructed image. For example, when $\psi_l(t) = t^2/2$, i.e., a quadratic potential function, the reconstructed images become smooth everywhere including the edges. Since this is undesirable, edge-preserving potential functions are used. These functions can reduce noise while preserving edges because they are quadratic near zero but linear above certain value, δ. The parameter δ determines the threshold for what values of contrast differences should be preserved. Here are some examples of edge-preserving potential functions:

- Hyperbola [94]:

$$\psi(t) = \delta^2 \left[\sqrt{1 + |t/\delta|^2} - 1 \right], \tag{2.18}$$

- q-Generalized Gaussian [124]:

$$\psi(t) = \frac{\frac{1}{2} |t|^p}{\left(1 + |t/\delta|^{(p-q)r} \right)^{1/r}}, \tag{2.19}$$

- Fair [35]

$$\psi(t) = \delta^2 \left[|t/\delta| - \log(1 + |t/\delta|) \right]. \tag{2.20}$$

We mostly focus on using convex and edge-preserving regularization in this thesis, and assume $\psi_l = \psi, \forall l$.

2.1.5 Iterative optimization algorithm

In statistical image reconstruction, the reconstructed image $\hat{\boldsymbol{x}}$ is obtained by minimizing the cost function (2.8). In general, there is no closed form solution for the minimizer $\hat{\boldsymbol{x}}$. Even for the quadratic cost function case where an analytical solution exists, direct calculation of this solution is impractical due to the enormous size of the system matrix. Therefore, we must use an iterative algorithm to minimize the cost function. One starts with an initial guess $\boldsymbol{x}_0$, which is typically the available FBP image, and then proceeds to refine the solution recursively. The reason why statistical methods are less practical compared to the conventional FBP approach is this necessity for an iterative process, which results in high computation time. Various methods have been proposed to obtain the solution in efficient manner, and thus reduce the computational burden [8, 74, 87, 139].

2.2 Motion-compensated image reconstruction (MCIR) methods

Even with the fast acquisition speed of commercial scanners, motion artifacts such as blurring and streaks are still a significant issue in CT image reconstruction, especially for cardiac CT imaging. Various methods have been proposed to address this problem [47, 50, 68, 109, 120]. Many of these are gated reconstruction methods that use only the projection data corresponding to approximately the same motion state, preferrably from a phase with the least cardiac motion. Diastolic phase of the heart cycle is an ordinary choice. For most practical purposes, a temporal resolution of 100 msec or less is enough to obtain reasonable images from diastole [111].

Gated reconstruction methods can be categorized into prospective ECG triggering and retrospective ECG gating. They both use simultaneously recorded ECG signals, but in different ways. Prospective ECG triggering uses a ECG signal to determine when to acquire data. On the other hand, retrospective ECG gating utilizes it to group the scan data from the same phase for image reconstruction. These methods can provide promising results in terms of image quality and processing time, but they suffer from limitations such as dose inefficiency and limited temporal resolution. Prospective ECG triggering is more dose efficient, but depends highly on a regular heart rate of the patient, which can lead to misregistration in the presence of arrhythmia. Retrospective ECG gating enables continuous and phase-consistent coverage of the entire heart, but suffers from inefficient dose [111]. Especially for fast and arrhythmic cardiac motion, both methods may be subjected to residual motion artifacts [63].

Motion-compensated image reconstruction (MCIR) methods exploit all collected data and use the motion information to obtain better spatial and temporal resolution and less motion-blurring [27,31,48,51,63,121–123,129]. This section of the chapter describes the mathematical models for the MCIR method, and they are general enough to represent different medical imaging modalities. However, specific assumptions will be made focusing on CT.

2.2.1 Measurement model

In addition to previously defined measurement model for CT (2.4), we add temporal dependence as follows. Let $\boldsymbol{x}(\boldsymbol{r}, t)$ denote the time-dependent attenuation coefficient distribution of the unknown object, where $\boldsymbol{r}$ is the spatial location and t is time. Let t_m be the time of mth frame at which the measurements, $\boldsymbol{y}_m$, corresponding to the motion-free state of the objects are acquired. We assume that the measurements consist of N_f scans, $\boldsymbol{y} = [\boldsymbol{y}_1, \cdots, \boldsymbol{y}_{N_f}]$. The measurements were assumed to be linearly related to the object $\boldsymbol{x}_m = \{x(\cdot, t_m)\}$ as follows:

$$\boldsymbol{y}_m = \boldsymbol{A}_m \boldsymbol{x}_m + \epsilon_m, \quad m = 1, \cdots, N_f, \tag{2.21}$$

where $\boldsymbol{A}_m$ is the system model for mth frame and ϵ_m is the noise. For the cases of CT that rotate during the scan or MRI that have dynamically changing sampling, the system model $\boldsymbol{A}_m$ will be different for each frame. On the other hand, for cases like a video sequence or an attenuation corrected PET, we can assume the same system model for every frame [26].

2.2.2 Deformation model

Throughout the thesis, we use cubic B-splines as our deformation models, because of their properties such as smoothness and computational efficiency including separability. A nonrigid transformation $\underline{T}$ in 3D can be represented as

$$\underline{T}(\boldsymbol{r}) = \boldsymbol{r} + \underline{d}(\boldsymbol{r}), \tag{2.22}$$

where $\boldsymbol{r} = (x, y, z)$ denotes the 3D Cartesian spatial coordinates, and $\underline{d}(\boldsymbol{r})$ is the displacement. A tensor product representation of the 3D displacement $\underline{d} = (d_x, d_y, d_z)$ is

$$d_p(\boldsymbol{r}; \underline{\alpha}) = \sum_{i,j,k} \alpha_{ijk}^p B_s\left(\frac{x}{m_x} - i\right) B_s\left(\frac{y}{m_y} - j\right) B_s\left(\frac{z}{m_z - k}\right), \tag{2.23}$$

where m_p is a spacing between knots in the p direction where $p \in \{x, y, z\}$, and B_s is a cubic B-spline basis function. Other deformation models like thin plate splines (TPS) [82] are also used.

2.2.3 Object and interpolation model

The object $x_m(\boldsymbol{r}) \triangleq x(\boldsymbol{r}, t_m)$ can be parameterized with a basis function B:

$$x_m(\boldsymbol{r}) = \sum_{l=1}^{N_b} c_{ml} B(\boldsymbol{r} - \boldsymbol{q}_l), \tag{2.24}$$

where N_b is the number of basis functions and $\boldsymbol{q}_l$ denotes the spatial coordinate of the center of the lth basis function. From (2.23) and (2.24), a deformed object can now be represented as

$$x_m(\boldsymbol{r} + \underline{d}(\boldsymbol{r}, \underline{\alpha})) = \sum_{l=1}^{N_b} c_{ml} B(\boldsymbol{r} + \underline{d}(\boldsymbol{r}; \underline{\alpha}) - \boldsymbol{q}_l). \tag{2.25}$$

In a simple matrix-vector form, (2.25) can be rewritten as follows:

$$x_m(\boldsymbol{r}_n + \underline{d}(\boldsymbol{r}_n, \underline{\alpha})) = \boldsymbol{T}(\underline{\alpha})\boldsymbol{c}_m, \tag{2.26}$$

where $\boldsymbol{x}_m = [x_m(\boldsymbol{r}_1), \cdots, x_m(\boldsymbol{r}_{n_\mathrm{p}})]$ denote a discretized vector of $x_m(\boldsymbol{r})$ with n_p samples located at $\{\boldsymbol{r}_n\}$, $\boldsymbol{c}_m = [c_{m1}, \cdots, c_{mN_b}]$ is the image coefficients for the mth frame image, and $\boldsymbol{T}(\underline{\alpha})$ is the $n_\mathrm{p} \times N_b$ warp matrix that has elements

$$[\boldsymbol{T}(\underline{\alpha})]_{nl} = B(\boldsymbol{r}_n + \underline{d}(\boldsymbol{r}_n; \underline{\alpha}) - \boldsymbol{q}_l).$$

In this thesis, we use a tensor-product of cubic B-splines as our basis function B.

2.2.4 Warp model

For convenience, we define the following notations:

$$\boldsymbol{x}_j = \boldsymbol{T}_{j,i}\boldsymbol{c}_i, \quad \boldsymbol{T}_0 = \boldsymbol{T}_{i,i}, \tag{2.27}$$

where $i, j = 1, \cdots, N_f$, $\boldsymbol{T}_{j,i}$ is a warp from the coefficients of ith frame to the jth frame image, and $\boldsymbol{T}_0$ denotes the interpolation operation with given coefficients. For our analysis, we use the following approximations for the image to image warp

$$\boldsymbol{x}_j \approx \boldsymbol{T}_{j,i}\boldsymbol{T}_0^{-1}\boldsymbol{x}_i, \tag{2.28}$$

where $\boldsymbol{T}_0^{-1}$ operation with B-spline basis functions is well-defined [128]. Finally, for notational simplicity, we denote $\boldsymbol{T}_i$ as the warp from a reference image $\boldsymbol{x}_\mathrm{ref}$ to ith frame image $\boldsymbol{x}_i$:

$$\boldsymbol{x}_i \approx \boldsymbol{T}_{i,\mathrm{ref}}\boldsymbol{T}_0^{-1}\boldsymbol{x}_\mathrm{ref} \triangleq \boldsymbol{T}_i\boldsymbol{x}_\mathrm{ref}. \tag{2.29}$$

2.2.5 Cost function formulation for MCIR

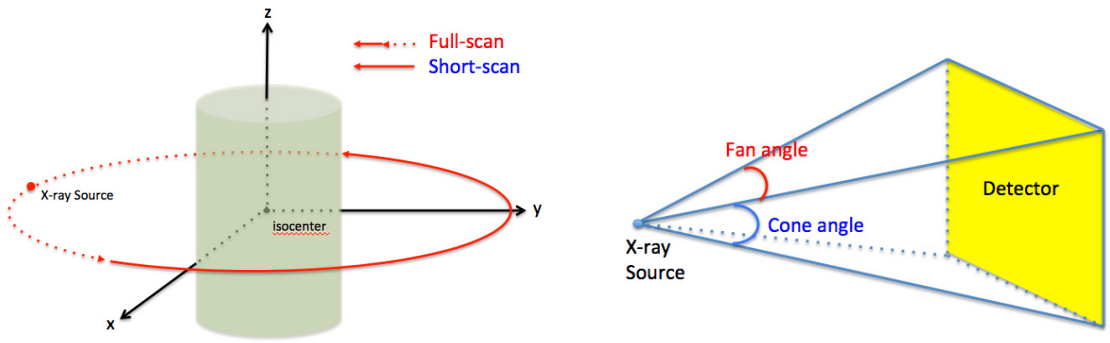
The cost function for MCIR methods can be represented in the following typical form [26]

$$\Psi(\boldsymbol{y}_1, \cdots, \boldsymbol{y}_{N_f} | \boldsymbol{x}_1, \cdots, \boldsymbol{x}_{N_f}, \underline{\alpha}_1, \cdots, \underline{\alpha}_{N_f}). \tag{2.30}$$

The goal is to reconstruct $\{\boldsymbol{x}_m\}$ from $\{\boldsymbol{y}_m\}$ using a motion model. The motion parameters $\{\underline{\alpha}_m\}$ may be estimated in advance or sequentially from separate motion estimation methods [11, 45, 55, 61, 64], or jointly estimated [13, 27, 122, 123]. The cost function (2.30) typically consists of a data-fidelity term, a regularizer for the reconstructed image, and a regularizer for the motion parameters for the case of joint estimation. The detailed formulations of the cost function will be described more in detail in Chapter VI.

2.3 Introduction to short-scan artifacts

In this section, we describe short-scan artifacts that appear in the reconstructed image of 3D cone-beam CT scans, especially when the short-scan protocol is used.



(a) (b)

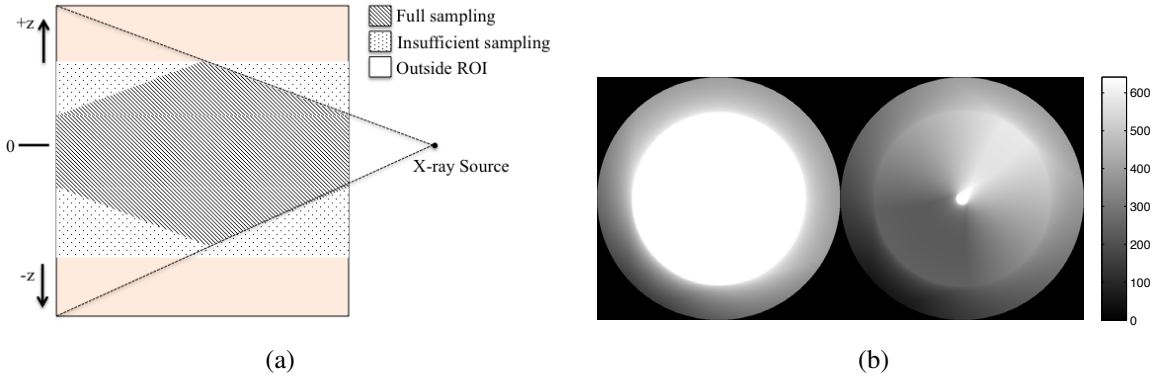
Figure 2.1: (a) X-ray source trajectories of full-scan and short-scan approaches for axial CT. (b) Fan and cone angles of X-ray cone-beam.

Due to significantly reduced data acquisition time, the short-scan approach, which has angular span of $\pi+2\gamma$ where γ is the fan angle of the projection (Fig. 2.1 (b)), has advantages like improved temporal resolution and lower dose compared to the full-scan approach at the expense of increased image noise level, unless the tube current is increased accordingly (Fig. 2.1 compares X-ray source trajectories of full-scan and short-scan approaches for axial CT). However, reduced angular span from 2π to $\pi + 2\gamma$ leads to insufficient sampling[3] at certain locations in third-generation CT geometry.

Fig. 2.2 shows that as we move away from the iso-plane more voxels are being insufficiently sampled, and thus end slices are most affected. Most of the voxels in end slices has sampling from

[3]In this study, "full" or "sufficient" sampling do not mean that they satisfy the complete sampling conditions derived in [92, 127], but rather mean that the voxel is seen in every projection view. Thus, insufficient sampling indicates the voxel is seen from only part of the projection views.

less than short-scan amount of views, and this situation will become more severe as the cone angle becomes larger (Fig. 2.1 (b)).



(a) (b)

Figure 2.2: (a) Diagram of sampling at each slice for axial CT geometry (b) Number of views passing through each voxel for 64 slice scanner with 70 cm FOV: Center slice (left) and end slice (right). A short-scan with 642 angular views were simulated and a circular mask was used. Display range is [0 642].

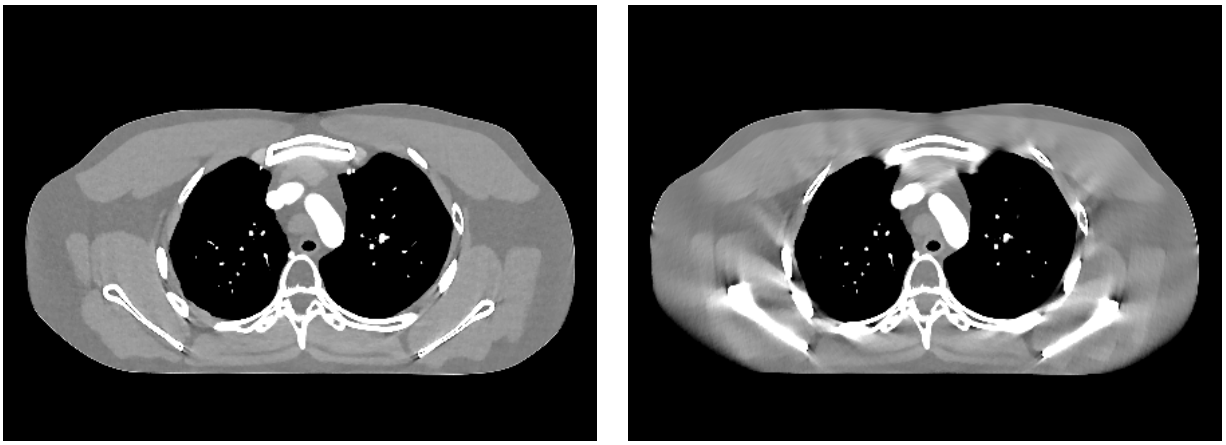
Insufficient sampling at end slices, when regularized iterative reconstruction method is used to reconstruct an image from short-scan projection data, can cause the following artifacts.

2.3.1 Distortion and shading

Fig. 2.3 compares the reconstructed images of the XCAT phantom at an end-of-ROI slice for both full-scan and short-scan measurements. A 256 slice CT scanner was simulated to obtain the measurements. The reconstructed image of short-scan data, on the contrary to that of full-scan data, shows severe distortions in some structures, especially around high contrast structures, and shadings at several locations.

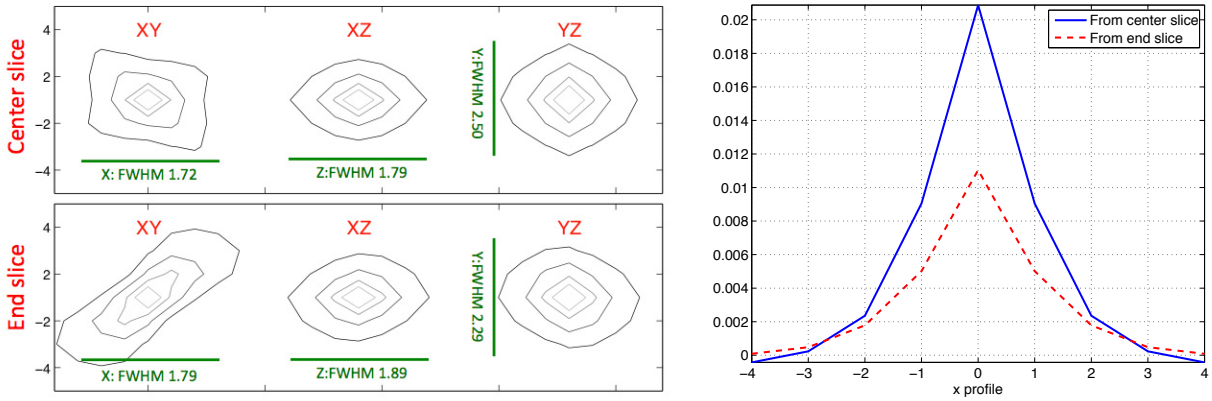
2.3.2 Resolution nonuniformity

Short-scan measurements corresponding to a 256-slice scanner were simulated to observe local impulse responses, defined as (3.6) in Chapter III, at different locations. Fig. 2.4 (a) shows contour plots of local impulse responses at two different locations from a center slice and an end slice, respectively. We can observe that each local impulse response has an anisotropic shape and also is different from each other. Such anisotropic shapes of local impulse responses represent anisotropic resolution in the reconstructed image (see Chapter III). As shown in Fig. 2.4(b), local impulse responses at different locations have different peak values, which will result in resolution nonuniformity in the reconstructed image (see Chapter IV).



(a) (b)

Figure 2.3: Reconstructed images of the XCAT phantom at end slice from (a) full-scan projection data (b) short-scan projection data. A 256 slice CT scanner was assumed to obtain the measurements.



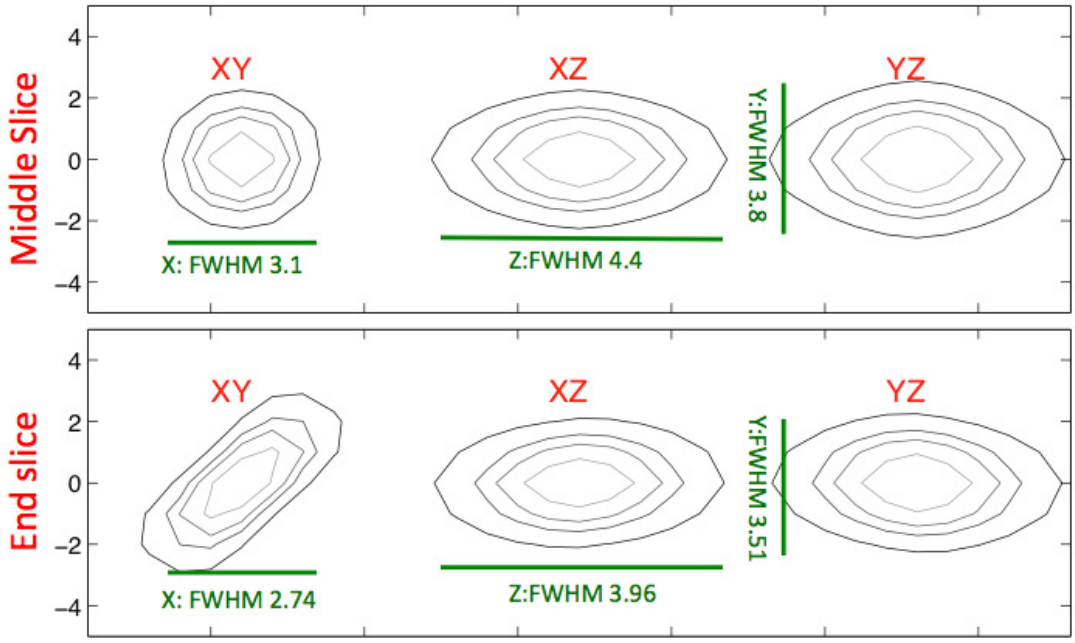
(a) (b)

Figure 2.4: Comparison of local impulse responses at two different locations: a center slice and an end slice. A 256-slice scanner with short-scan measurements were simulated. (a) Contour plots of xy, xz, and yz profiles. Each contour was plotted based on its own peak value. (b) Plot of x profiles through the center of each impulse response.

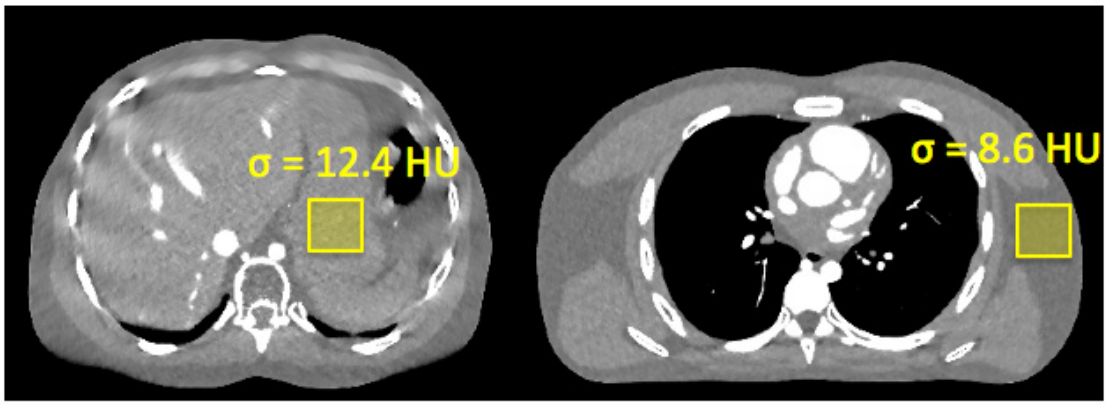
2.3.3 Nonuniform and correlated noise characteristic

Fig. 2.5 (a) compares the local noise correlations, defined as (3.37) in Chapter III, at two different slices: a middle slice and an end slice. A 256-slice CT scanner was simulated and the XCAT phantom was used as an object. The local noise correlations show nonuniform and anisotropic shapes that will result in nonuniform and correlated noise property in the reconstructed image. Reconstructed images of the XCAT phantom at an end slice and a middle slice are shown in Fig. 2.5 (b). Standard deviation of the image values at the selected uniform regions shows that the noise is

nonuniform in the reconstructed image.



(a)



(b)

Figure 2.5: (a) Comparison of local noise correlations at two different locations: a middle slice and an end slice. Contour plots of xy, xz, and yz profiles are presented. Each contour was plotted based on its peak value. (b) Reconstructed images of the XCAT phantom at an end slice (left) and a middle slice (right). Standard deviation of image values at the selected regions are shown.

Even though the short-scan artifacts were categorized as above, they are closely interconnected to each other and appear simultaneously in the reconstructed image. These artifacts are undesirable and may prohibit accurate diagnosis, and thus need to be removed or at least mitigated.

In the following chapters, we investigated various methods that addresses different short-scan artifacts. Two Fourier domain based regularization design methods in Chapter III address anisotropic and nonuniform spatial resolution and nonuniform and correlated noise property, respectively. In Chapter IV, regularization design methods using a hypothetical geometry concept focuses on improving uniformity of spatial resolution or noise in the reconstructed image. Finally, Chapter V presents methods that remove most of short-scan artifacts, especially distortions and shadings, by using all available information.

CHAPTER III

Regularization designs for isotropic and uniform spatial resolution and noise

While statistical image reconstruction methods have potential advantages over conventional FBP reconstruction such as reduced patient dose and improved noise properties, their use of statistical weighting and space variant scanning geometries can lead to nonuniform and anisotropic spatial resolution and noise characteristics. This holds even for idealized shift-invariant imaging systems [42], and becomes most severe near the end slices of 3D axial or helical CT.

Conventional regularizers defined as (2.17) have space variant nature, but only in the sense of overall regularization strength at each location. Regularization strength between a voxel and its neighbors is usually determined by object independent factors such as voxel size and distance between a voxel and its neighbor, and thus is the same for all voxels. Such regularizers are not enough to compensate for anisotropic image properties induced by statistical weighting and space variant scanning geometries. Therefore, in this chapter, we investigate regularizers having spatially varying regularization strength between each voxel and its neighbors.

As a preliminary study, we investigated heuristic regularizer designs for mitigating short-scan artifacts, especially nonuniform spatial resolution. Our goal for this study was to learn whether modifying the directional strength of the regularizer can lead to improved resolution uniformity and reduced distortion. The basic idea to overcome the lack of sampling was to make the voxels with insufficient sampling depend more on the neighbors with better sampling. Motivated by the success of zeroth-order extrapolation for FBP reconstruction, we investigated a design that enforces stronger axial regularization in axial direction and less for other directions. Having strong axial direction regularization for insufficiently sampled region may have similar effect as what FBP is doing in practice. Experiments with clinical and phantom data showed some promise for improving resolution uniformity. However, the methods showed clear limitations including dependence on empirical parameter selection. Details of this preliminary study are not presented

This chapter is based on [19, 20]

in the thesis, but the study motivated more systematic regularization design methods presented in this Chapter.

We propose quadratic regularization design methods for 3D axial X-ray computed tomography (CT) that aim to improve isotropy and uniformity of resolution and noise, respectively. Simulations and a phantom experiment show that the proposed methods lead to more uniform and isotropic spatial resolution and noise characteristics in 3D axial CT with modest computational cost.

The focus of these methods is to design the regularization parameter, β_j^l, to achieve the goal. Instead of directly modifying the regularization parameter β_j^l in (2.14), we design the directional weights $\{r^l\}$ in (2.17) for each voxel, leading to $\{r_j^l\}$. The main reason for taking this approach is that the directional weights $\{r^l\}$ have much smaller dynamic range compared to the regularization parameter β_j^l and can be more intuitively constrained by certain conditions, such as fixed norm, during the design process. However, the designs proposed in this chapter can be readily applied to directly modifying the regularization parameter β_j^l.

3.1 Quadratic regularization design for isotropic and uniform spatial resolution

Some may argue that forcing isotropic resolution can interrupt detection and quantification of lesions [96]; however the importance and consequences of isotropic and uniform spatial resolution still require more investigation. In addition, the method developed in this section is not limited to isotropic and uniform spatial resolution designs, but can be used for other user-defined spatial resolution criteria. The spatial resolution of the reconstructed image should be controlled by the users, and not just obtained as an arbitrary byproduct of the estimator used [113].

Several previous regularization designs aim to match the local impulse response of the estimator to a target impulse response by matrix manipulations and discrete Fourier transforms [42, 119]. Due to the large number of voxels in the image volume, regularization design methods based on discrete Fourier transforms would require prohibitive computational cost, especially for 3D axial or helical CT. A fast analytical regularization design method for 2D fan-beam X-ray CT that uses continuous space analogs to simplify the regularization design problem was proposed in [116]. In [115], the authors addressed the problem for 3D axial CT, but for a simplified 3D system that was modeled as a stack of 2D fan-beam systems. In this section, we propose a regularization design for 3D axial X-ray computed tomography (CT) accounting for cone angle. Simulations and a phantom experiment show that the proposed method leads to more uniform and isotropic spatial resolution in 3D axial CT with modest computational cost.

3.1.1 Local frequency response

From (2.9), (2.10), and (2.12), we obtain the following penalized weighted least squares (PWLS) objective function

$$\Psi(\boldsymbol{x}) = \text{Ł}(\boldsymbol{x}) + \mathsf{R}(\boldsymbol{x}), \quad \text{Ł}(\boldsymbol{x}) = \frac{1}{2} \left\| \boldsymbol{y} - \boldsymbol{A}\boldsymbol{x} \right\|_{\boldsymbol{W}}^2, \tag{3.1}$$

where $\boldsymbol{W} = \text{diag}\{w_i\}$ is a statistical weighting matrix, and $\mathsf{R}(\boldsymbol{x})$ is defined as (2.17). Our purpose is to improve the regularizer $\mathsf{R}(\boldsymbol{x})$, by designing the directional weights $\{r_j^l\}$, such that our local impulse response (LIR) $\boldsymbol{l}_j$ matches a target l^0 that has more isotropic spatial resolution, at every voxel j. We simplify this process by turning to the frequency domain. As mentioned in the beginning of the chapter, the design process in the section can be readily applied to directly modifying the regularization parameter β_j^l in (2.14).

When we use different weights for each location or asymmetric weights, our iterative algorithm also needs to be modified to take that into account. The gradient of the regularizer (2.17) now becomes:

$$\begin{aligned}
\nabla_{\boldsymbol{x}} \mathsf{R}(\boldsymbol{x}) &= \left[\cdots, \frac{\partial}{\partial x_j} \mathsf{R}(\boldsymbol{x}), \cdots \right]', && (3.2) \\
\frac{\partial}{\partial x_j} \mathsf{R}(\boldsymbol{x}) &= \beta \kappa_j \sum_{l \in N_j} r_j^l \kappa_{j-j_l} \dot{\psi}(x_{j-j_l}, x_j) + \beta \sum_{k: j \in N_k} r_k^{l_k(j)} \kappa_k \kappa_j \dot{\psi}(x_j, x_k) && (3.3) \\
&= \beta \kappa_j \sum_{l \in N_j} \left(r_j^l + r_{j-j_l}^{l_{j-j_l}(j)} \right) \kappa_{j-j_l} \dot{\psi}(x_j, x_{j-j_l}), && (3.4)
\end{aligned}$$

where $l_k(j)$ means the lth directional weight of a voxel located at k that is linked to the voxel located at j. Notice that the weight between two voxels appears as the sum of the weights to each other.

We first define the local impulse response. The local impulse response describes the local spatial resolution properties. The following definition of local impulse response at jth voxel was used [116]:

$$\begin{aligned}
\boldsymbol{l}_j &\triangleq \lim_{\epsilon \to 0} \frac{\hat{\boldsymbol{x}}\left(\bar{\boldsymbol{y}}\left(\boldsymbol{x}^{\text{true}} + \epsilon \boldsymbol{\delta}_j\right)\right) - \hat{\boldsymbol{x}}\left(\bar{\boldsymbol{y}}\left(\boldsymbol{x}^{\text{true}}\right)\right)}{\epsilon} && (3.5) \\
&= \nabla_{\boldsymbol{y}} \hat{\boldsymbol{x}}\left(\boldsymbol{y}\right) |_{\boldsymbol{y} = \bar{\boldsymbol{y}}(\boldsymbol{x}^{\text{true}})} \nabla_{\boldsymbol{x}} \bar{\boldsymbol{y}}\left(\boldsymbol{x}\right) |_{\boldsymbol{x} = \boldsymbol{x}^{\text{true}}} \boldsymbol{\delta}_j, && (3.6)
\end{aligned}$$

where $\boldsymbol{\delta}_j$ is a Kronecker impulse at jth voxel, and the gradient operations are matrices with the following elements:

$$\left[\nabla_{\boldsymbol{y}} \hat{\boldsymbol{x}}\left(\boldsymbol{y}\right)\right]_{ji} = \frac{\partial}{\partial y_i} \hat{x}_j(\boldsymbol{y}), \quad \left[\nabla_{\boldsymbol{x}} \bar{\boldsymbol{y}}\left(\boldsymbol{x}\right)\right]_{ij} = \frac{\partial}{\partial x_j} \bar{y}_i(\boldsymbol{x}).$$

The impulse response given above is local in the sense that it is shift-variant and depends on the object $\boldsymbol{x}^{\text{true}}$ at the voxel.

For simplicity we focus on quadratic regularization, for which, from (3.6), the local impulse response of the PWLS estimator (2.8) of the cost function (3.1) is expressed as

$$\boldsymbol{l}_j = [\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A} + \boldsymbol{R}]^{-1}\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A}\boldsymbol{\delta}_j, \tag{3.7}$$

where $\boldsymbol{R}$ is the Hessian of the regularizer $\mathsf{R}(\boldsymbol{x})$ [42]. The statistical weighting $\boldsymbol{W}$ depends on $\boldsymbol{x}^{\text{true}}$ due to Poisson-type statistics.

Assuming $\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A}\boldsymbol{\delta}_j$ and $\boldsymbol{R}\boldsymbol{\delta}_j$ are approximately locally circulant [99], we can approximate (3.7) as follows:

$$L^j = \frac{F\left(\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A}\boldsymbol{\delta}_j\right)}{F\left(\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A}\boldsymbol{\delta}_j\right) + \beta F\left(\boldsymbol{R}\boldsymbol{\delta}_j\right)}, \tag{3.8}$$

where $F(\cdot)$ denotes the 3-D DFT centered at voxel j.

Instead of directly using the discrete Fourier transform, we use the continuous-space analog of $H_j \triangleq F\left(\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A}\boldsymbol{\delta}_j\right)$ in spherical coordinates $\boldsymbol{\nu} \triangleq (\rho, \Phi, \Theta)$. We use a closed-form approximation for H_j that was suggested in [110]:

$$\begin{aligned}
H_j(\boldsymbol{\nu}) &\approx KJ(\boldsymbol{\nu})\frac{\tilde{w}_j(\Phi)}{\rho\cos(\Theta)} \\
K &= \Pi\Delta_x^3\Delta_z D_{sd}^2/D_{so}^2 \\
J(\boldsymbol{\nu}) &= \operatorname{sinc}(\Delta_x\rho\cos(\Theta)\cos\Phi)^2 \\
&\quad \times\operatorname{sinc}(\Delta_y\rho\cos(\Theta)\sin\Phi)^2\operatorname{sinc}(\Delta_z\rho\sin(\Theta))^2 \\
\tilde{w}_j(\Phi) &= \sum_{\beta\in B_j(\Phi)} \frac{\bar{w}_{\beta,j}}{d_{\beta,j}\sqrt{1-(\zeta^j\cos(\theta^j))^2\cos^2(\phi^j-\Phi)}},
\end{aligned} \tag{3.9}$$

where $D_{so}\cdot(\zeta^j, \phi^j, \theta^j)$ denotes the location of the jth voxel in spherical coordinates, K is a constant depending on voxel sizes and scanner geometry, $J(\boldsymbol{\nu})$ is a factor depending only on spatial frequencies, $\bar{w}_{\beta,j} \triangleq w_\beta(\bar{s}_j^*)$ where $\bar{s}_j^*$ is the position on the detector that maximizes the footprint of voxel j at source angle β, $d_{\beta,j}$ is the distance from the source to the xy-projection of voxel j, and $B_j(\Phi)$ is the set of the values of β for which the ray passing through voxel j is perpendicular to the frequency vector $\boldsymbol{\nu}$ where the ray and frequency vector are both projected onto the xy-plane [110]. Substituting (3.10) into (3.8) yields the following expression for the continuous space analog of L^j:

$$L^j \approx \frac{KJ(\boldsymbol{\nu})\tilde{w}_j(\Phi)/(\rho\cos(\Theta))}{KJ(\boldsymbol{\nu})\tilde{w}_j(\Phi)/(\rho\cos(\Theta)) + \beta R_j(\boldsymbol{\nu})}, \tag{3.10}$$

where $R_j(\boldsymbol{\nu})$ is the local frequency response for the regularizer near pixel j (see (3.18) below).

3.1.2 Target frequency response

The local frequency response associated with penalized unweighted reconstruction is isotropic at the isocenter for a full-scan, so we use it as our target response. At isocenter, (3.10) for uniform weights ($w_i = 1$) is given as

$$\begin{aligned} H_0(\boldsymbol{\nu}) &\approx KJ(\boldsymbol{\nu})\frac{\tilde{u}_0(\Phi)}{\rho\cos(\Theta)}, \\ \tilde{u}_0(\Phi) &= |B_j(\Phi)|\,. \end{aligned} \tag{3.11}$$

Now the target local frequency response is

$$L^0 \approx \frac{KJ(\boldsymbol{\nu})\tilde{u}_0(\Phi)/(\rho\cos(\Theta))}{KJ(\boldsymbol{\nu})\tilde{u}_0(\Phi)/(\rho\cos(\Theta)) + \beta R_0(\boldsymbol{\nu})}, \tag{3.12}$$

where L^0 is the continuous-space analog of L^0.

Our purpose is to match the local impulse response at jth voxel to the target impulse response, i.e., we want

$$\begin{aligned} L^j &\approx \frac{KJ(\boldsymbol{\nu})\tilde{w}_j(\Phi)/(\rho\cos(\Theta))}{KJ(\boldsymbol{\nu})\tilde{w}_j(\Phi)/(\rho\cos(\Theta)) + \beta R_j(\boldsymbol{\nu})} \\ &\approx \frac{KJ(\boldsymbol{\nu})\tilde{u}_0(\Phi)/(\rho\cos(\Theta))}{KJ(\boldsymbol{\nu})\tilde{u}_0(\Phi)/(\rho\cos(\Theta)) + \beta R_0(\boldsymbol{\nu})} \approx L^0. \end{aligned} \tag{3.13}$$

Cross multiplying and simplifying yields the goal

$$\tilde{u}_0(\Phi)R_j(\boldsymbol{\nu}) \approx \tilde{w}_j(\Phi)R_0(\boldsymbol{\nu}). \tag{3.14}$$

3.1.3 Regularization structure

Taking the Fourier transform of (2.15) yields the following expression for the local frequency response

$$\begin{aligned} |C_l(\omega_1,\omega_2,\omega_3)|^2 &= \frac{1}{n_l^2+m_l^2+z_l^2}\left|1-e^{-i(\omega_1 n_l+\omega_2 m_l+\omega_3 z_l)}\right|^2 \\ &= \frac{1}{n_l^2+m_l^2+z_l^2}\left(2-2\cos\left(\omega_1 n_l+\omega_2 m_l+\omega_3 z_l\right)\right). \end{aligned} \tag{3.15}$$

Using the approximation $2-2\cos(x) \approx x^2$ [116] (3.15) simplifies

$$|C_l(\omega_1,\omega_2,\omega_3)|^2 \approx \frac{1}{n_l^2+m_l^2+z_l^2}\left(\omega_1 n_l+\omega_2 m_l+\omega_3 z_l\right)^2. \tag{3.16}$$

We convert (3.16) to spherical frequency coordinates. The relationship between frequency and sampling yields $\omega_1 = 2\pi\Delta_x\rho\cos(\Phi)\cos(\Theta)$, $\omega_2 = 2\pi\Delta_y\rho\sin(\Phi)\cos(\Theta)$, and $\omega_3 = 2\pi\Delta_z\rho\sin(\Theta)$. Substituting these into (3.16) yields the following expression for

$$\begin{aligned}&|C_l(\omega_1,\omega_2,\omega_3)|^2\\ &\approx \frac{1}{n_l^2+m_l^2+z_l^2}(2\pi\rho)^2(n_l\Delta_x\cos(\Phi)\cos(\Theta)+m_l\Delta_y\sin(\Phi)\cos(\Theta)+z_l\Delta_z\sin(\Theta))^2.\end{aligned} \tag{3.17}$$

The local frequency response of the regularizer (2.17) is now

$$R_j(\rho,\Phi,\Theta) = (2\pi\rho)^2\kappa_j^2\sum_{l=1}^{N_l} r_j^l\left(e(\Phi,\Theta)\cdot[e(\Phi_l,\Theta_l)\otimes\Delta]\right)^2, \tag{3.18}$$

where $e(\Phi,\Theta) \triangleq (\cos(\Phi)\cos(\Theta),\sin(\Phi)\cos(\Theta),\sin(\Theta))$, $\Delta \triangleq (\Delta_x,\Delta_y,\Delta_z)$, $\otimes$ is element-wise multiplication, and we assumed that $\kappa_j \approx \kappa_l$ for l within the neighborhood of j.

For the target response, R_0 becomes

$$R_0(\rho,\Phi,\Theta) = (2\pi\rho)^2\kappa_0^2\sum_{l=1}^{N_l} r_0^l\left(e(\Phi,\Theta)\cdot[e(\Phi_l,\Theta_l)\otimes\Delta]\right)^2, \tag{3.19}$$

where κ_0 is the user-defined weights for target spatial resolution at the isocenter, and $\{r_0^l\}$ is the pre-defined directional weights, which determines the shape of the target response.

3.1.4 Regularization design for isotropic and uniform spatial resolution

Substituting (3.18) and (3.19) into (3.14) and simplifying yields

$$Q_j(\Phi,\Theta) \approx \frac{\kappa_0^2\tilde{w}_j(\Phi)}{\kappa_j^2\tilde{u}_0(\Phi)}Q_0(\Phi,\Theta), \tag{3.20}$$

where

$$Q_j(\Phi,\Theta) \triangleq \sum_{l=1}^{N_l} r_j^l\left(e(\Phi,\Theta)\cdot[e(\Phi_l,\Theta_l)\otimes\Delta]\right)^2, \tag{3.21}$$

$$Q_0(\Phi,\Theta) \triangleq \sum_{l=1}^{N_l} r_0^l\left(e(\Phi,\Theta)\cdot[e(\Phi_l,\Theta_l)\otimes\Delta]\right)^2. \tag{3.22}$$

We solve the following weighted minimization problem to design the directional weighting

coefficient vector $\boldsymbol{r}^j = (r_j^1, \cdots, r_j^{N_l})$ at the jth voxel

$$\boldsymbol{r}^j \triangleq \underset{r^j \geq 0}{arg\,min} \int_0^{2\pi} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} D_w(\Phi, \Theta) \left| \breve{w}_j(\Phi, \Theta) - \sum_{l=1}^{N_l} r_j^l (e(\Phi, \Theta) \cdot [e(\Phi_l, \Theta_l) \otimes \Delta])^2 \right|^2 d\Theta d\Phi, \tag{3.23}$$

where $D_w(\Phi, \Theta)$ is a user-defined weighting function, the non-negativity constraint ensures the regularizer's convexity and we define the modified weighting function

$$\breve{w}_j(\Phi, \Theta) \triangleq \frac{\kappa_0^2 \tilde{w}_j(\Phi)}{\kappa_j^2 \tilde{u}_0(\Phi)} \sum_{l=1}^{N_l} r_0^l (e(\Phi, \Theta) \cdot [e(\Phi_l, \Theta_l) \otimes \Delta])^2. \tag{3.24}$$

The choice of the weighting function can be determined by factors like the coordinate system. We choose $D_w = \cos(\Theta)$ to have more uniform effect of the sample points. We view (3.23) as a weighted projection of $\breve{w}_j(\Phi)$ onto the space spanned by $\{(e(\Phi, \Theta) \cdot [e(\Phi_l, \Theta_l) \otimes \Delta])^2\}$, which can be expanded as follows

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

where $(\Phi, \Theta) \in [0, 2\pi] \times [-\pi/2, \pi/2]$. From this equation, we can extract 6 basis functions as follows

$$\begin{aligned}
q_1(\Phi, \Theta) &= 1 \\
q_2(\Phi, \Theta) &= \cos^2(\Theta) \cos(2\Phi) \\
q_3(\Phi, \Theta) &= \cos^2(\Theta) \sin(2\Phi) \\
q_4(\Phi, \Theta) &= \cos(2\Theta) \\
q_5(\Phi, \Theta) &= \sin(2\Theta) \cos(\Phi) \\
q_6(\Phi, \Theta) &= \sin(2\Theta) \sin(\Phi),
\end{aligned}$$

where we used the following trigonometric identities

$$\begin{aligned}\sin^2(\Phi) &= \frac{1}{2}\left(1-\cos(2\Phi)\right)\\ \cos^2(\Phi) &= \frac{1}{2}\left(\cos(2\Phi)-1\right).\end{aligned}$$

Notice that regardless of the size of the neighborhood, the local frequency response of the regularizer (3.18) is spanned by only 6 functions. This limited degree of freedom suggests that we can only obtain modest variations for the local frequency response of the regularizer.

To obtain orthonormal basis functions, we define an inner product for (3.23)

$$\langle A, B\rangle = \frac{1}{2\pi^2}\int_0^{2\pi}\int_{-\pi/2}^{\pi/2}\cos(\Theta)\,A(\Theta,\Phi)B^*(\Theta,\Phi)d\Theta d\Phi. \tag{3.25}$$

The orthonormal basis functions are given as follows

$$\begin{aligned}p_1(\Phi,\Theta) &= \sqrt{\frac{\pi}{2}}\\ p_2(\Phi,\Theta) &= \sqrt{\frac{15\pi}{8}}\cos^2(\Theta)\cos(2\Phi)\\ p_3(\Phi,\Theta) &= \sqrt{\frac{15\pi}{8}}\cos^2(\Theta)\sin(2\Phi)\\ p_4(\Phi,\Theta) &= \sqrt{\frac{45\pi}{32}}\left(\cos(2\Theta)-\frac{1}{3}\right)\\ p_5(\Phi,\Theta) &= \sqrt{\frac{15\pi}{8}}\sin(2\Theta)\cos(\Phi)\\ p_6(\Phi,\Theta) &= \sqrt{\frac{15\pi}{8}}\sin(2\Theta)\sin(\Phi),\end{aligned}$$

where we used the Gram-Schmidt process to obtain p_4 because q_1 and q_4 were not orthogonal.

We can decompose $\sum_{l=1}^{N_l} r_j^l(e(\Phi,\Theta)\cdot[e(\Phi_l,\Theta_l)\otimes\Delta])^2$ as $\boldsymbol{PTr}^j$, where $\boldsymbol{P}$ is an operator whose columns are the six orthonormal vectors, and $\boldsymbol{T}$ is a $6\times N_l$ linear combination matrix whose mth row is the following inner product

$$\boldsymbol{T}_{ml} = \frac{1}{2\pi^2}\int_0^{2\pi}\int_{-\pi/2}^{\pi/2}\cos(\Theta)(e(\Phi,\Theta)\cdot[e(\Phi_l,\Theta_l)\otimes\Delta])^2 p_m d\Theta d\Phi. \tag{3.26}$$

Assuming $\Delta_x = \Delta_y$, the lth column of $\boldsymbol{T}$ is given by

$$\begin{bmatrix} \frac{2}{3\pi}\left[\Delta_x^2\cos^2(\Phi_l) + \Delta_z^2\sin^2(\Theta_l)\sin^2(\Phi_l)\right] \\ \frac{4}{15\pi}\Delta_x^2\cos^2(\Phi_l) \\ \frac{8}{15\pi}\Delta_x^2\cos^2(\Theta_l)\cos(\Phi_l)\sin(\Phi_l) \\ \frac{8}{45\pi}\left[\Delta_x^2\cos^2(\Phi_l) - 2\Delta_z^2\sin^2(\Theta_l)\sin^2(\Phi_l)\right] \\ \frac{8}{15\pi}\Delta_x\Delta_z\cos(\Theta_l)\cos(\Phi_l)\sin(\Theta_l) \\ \frac{8}{15\pi}\Delta_x\Delta_z\cos(\Theta_l)\sin(\Phi_l)\sin(\Theta_l) \end{bmatrix}.$$

Since $\boldsymbol{P}$ has orthonormal columns, we can represent the minimization problem (3.23) as the following simplified expression

$$\boldsymbol{r}^j = \underset{\boldsymbol{r}\geq 0}{arg\,min} \left\|\boldsymbol{T}\boldsymbol{r} - \boldsymbol{b}^j\right\|^2, \tag{3.27}$$

where $\boldsymbol{P}^*$ denotes the adjoint of $\boldsymbol{P}$ and $\boldsymbol{b}^j \triangleq \boldsymbol{P}^*\check{w}_j(\cdot)$, *i.e.*,

$$b_k^j = \frac{1}{2\pi^2}\int_0^{2\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos(\Theta)\, p_k(\Phi,\Theta)\check{w}_j(\Phi,\Theta)d\Phi d\Theta, \quad k = 1,\cdots,6.$$

The minimization problem (3.27) is much smaller than (3.23). For 3D case, solving (3.27) analytically would be not be trivial, instead, we solve (3.27) using NNLS algorithm [78]. Since (3.27) is a small problem with typical value of N_l equal to 13, only modest computation is required.

The minimization problem (3.27) is under-determined and may have many different solutions $\boldsymbol{r}^j$ that are all global minima. Too many zeros in $\boldsymbol{r}^j$ may degrade the image since there will zeros in the Hessian [116]. To ensure that certain $\boldsymbol{r}^j$ values are greater than some small positive number $\boldsymbol{\epsilon}_l^j$, we modify (3.27) as follows

$$\begin{aligned} \boldsymbol{r}^j &= \underset{\boldsymbol{r}\geq 0}{arg\,min} \left\|\boldsymbol{T}\boldsymbol{r} - (\boldsymbol{b}^j - \boldsymbol{T}\boldsymbol{\epsilon}^j)\right\|^2 &(3.28)\\ &= \underset{\boldsymbol{r}\geq 0}{arg\,min} \left\|\boldsymbol{T}\boldsymbol{r} - \bar{\boldsymbol{b}}^j\right\|^2, &(3.29)\end{aligned}$$

where $\bar{\boldsymbol{b}}^j \triangleq \boldsymbol{b}^j - \boldsymbol{T}\boldsymbol{\epsilon}^j$. After minimization, we use the coefficients $\bar{\boldsymbol{r}}^j = \boldsymbol{r}^j + \boldsymbol{\epsilon}^j$ for our new regularizer (See [116] for a possible way to select $\boldsymbol{\epsilon}^j$).

We also need the mapping from $\boldsymbol{P}^*\check{w}_j$ to $\boldsymbol{r}^j$ to be continuous for our locally circulant approximations. We use Tikhonov regularization to coerce our NNLS problem to select a solution that also minimizes the norm of $\boldsymbol{r}^j$ [114]. By appending a scaled identity matrix to $\boldsymbol{T}$ and zero padding

$\bar{\boldsymbol{b}}^j$, (3.29) can be further modified as follows

$$\boldsymbol{r}^j = \underset{\boldsymbol{r}\geq 0}{arg\,min} \left\| \bar{\boldsymbol{T}}\boldsymbol{r} - \hat{\boldsymbol{b}}^j \right\|^2 \tag{3.30}$$

$$= \underset{\boldsymbol{r}\geq 0}{arg\,min} \left\| \boldsymbol{T}\boldsymbol{r} - \bar{\boldsymbol{b}}^j \right\|^2 + \tau \left\| \boldsymbol{r} \right\|^2, \tag{3.31}$$

where

$$\bar{\boldsymbol{T}} = \begin{pmatrix} \boldsymbol{T} \\ \sqrt{\tau}\boldsymbol{I} \end{pmatrix}, \hat{\boldsymbol{b}} = \begin{pmatrix} \bar{\boldsymbol{b}} \\ \boldsymbol{0} \end{pmatrix}. \tag{3.32}$$

We select the value of τ to be a small number that is sufficient to apply a small penalty to the norm of $\boldsymbol{r}^j$.

3.1.5 XCAT simulation results

We simulated a 3rd-generation axial cone-beam CT system using the separable footprint projector [79]. The simulated system has $N_s = 888$ channels and $N_t = 64$ detector rows spaced by $\Delta_s = 1.0239$ mm and $\Delta_t = 1.09878$ mm, and 984 evenly spaced view angles over a 360 degree scan. The source to detector distance was 949 mm, and the source to rotation center distance was 541 mm. We included a quarter detector offset in the channel direction to reduce aliasing. The XCAT phantom [112] was used, and the image was reconstructed to a $512 \times 512 \times 122$ grid with pixel size $\Delta_x = \Delta_y = 0.9766$ mm and $\Delta_z = 0.625$ mm. Poisson noise was added to the sinogram, and the statistical weighting was chosen as $w_i = \exp(-[\boldsymbol{A}\boldsymbol{x}]_i)$. The regularization parameter β was selected such that the target PSF at the isocenter has a full-width at half-maximum (FWHM) of approximately 1.4 mm in xy and 0.9 mm in z.

Figs. 3.1 and 3.2 compare impulse responses of conventional regularization and proposed regularization for two different voxels with different sampling properties. There is a considerable anisotropy at both locations, especially for voxels with insufficient sampling. The main reasons for the anisotropy are statistical weighting and scan geometry. The spatial resolution of the voxel in Fig. 3.1 is primarily affected by the statistical weights, and our proposed method gives more isotropic impulse response. The location in Fig. 3.2 is greatly affected by scan geometry, and our proposed method achieves limited improvements.

Fig. 3.3 compares reconstructed images with various methods. Iteratively reconstructed images show better noise characteristics compared to the FDK reconstruction, but they may have more anisotropic spatial resolution especially at the voxels with less samplings. The true image blurred by the target impulse response was provided as a reference to assess the improvements of our proposed method. In Fig. 3.4 closely compares the reconstructed images with conventional regularization and the proposed regularization. Overall, the reconstructed image with the proposed

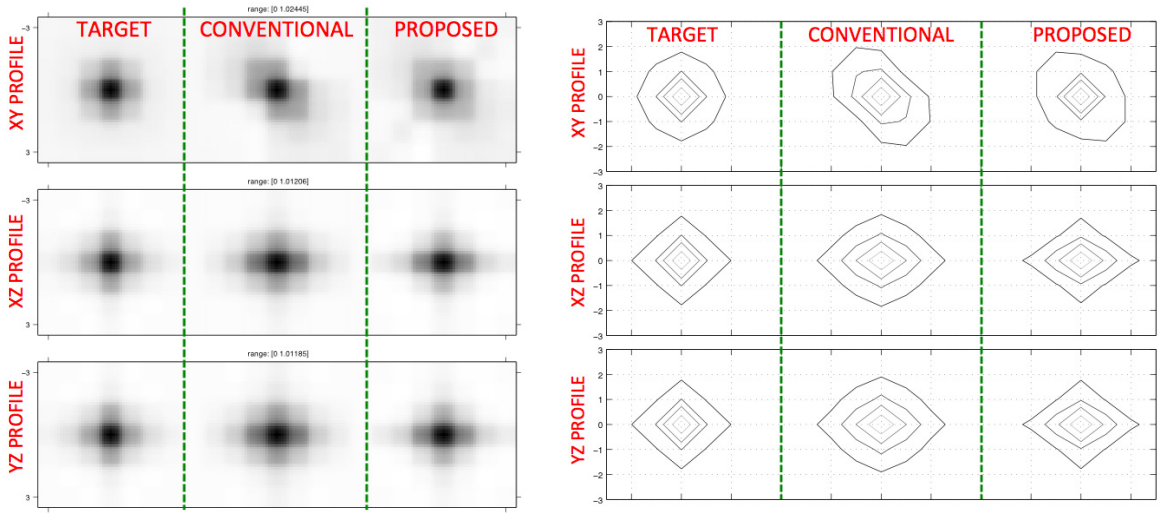


Figure 3.1: Impulse responses of conventional regularization (middle column) and proposed regularization (right column) at (-66,217,-17) (mm), which is a fully sampled location. Target impulse response is given as a reference (left column). Each row corresponds to xy, xz, and yz profiles, respectively. Each contour was plotted based on its own peak value.

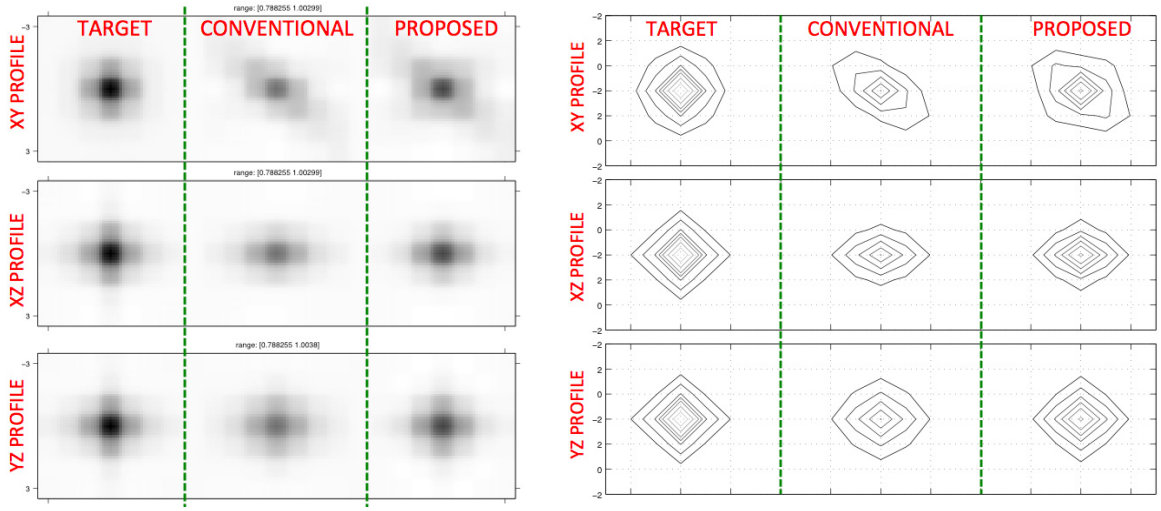
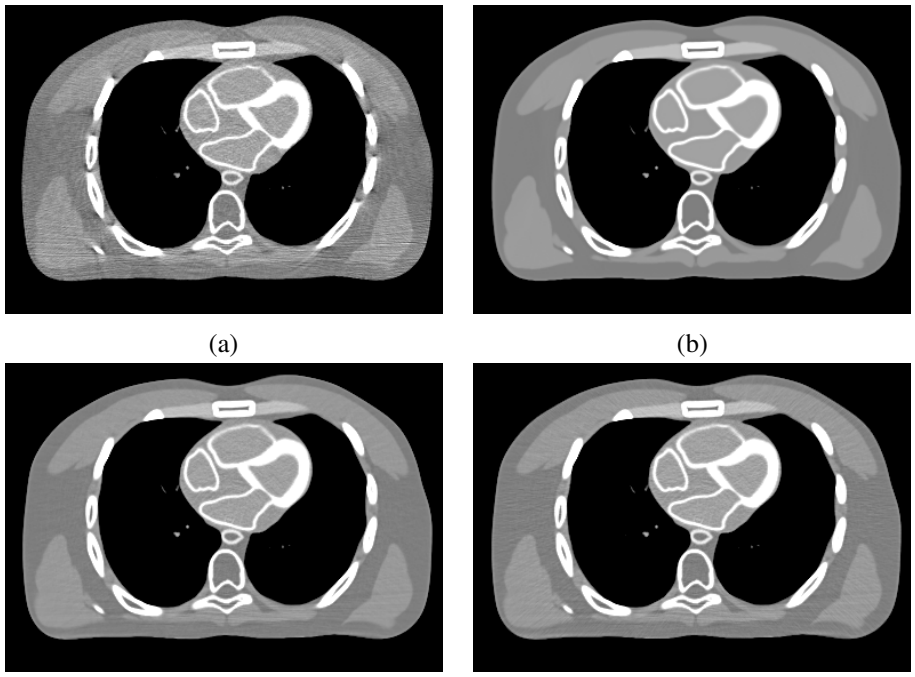


Figure 3.2: Impulse responses of conventional regularization (middle column) and proposed regularization (right column) at (-117,-67,17) (mm), which is a insufficiently sampled location. Target impulse response is given as a reference (left column). Each row corresponds to xy, xz, and yz profiles, respectively. Each contour was plotted based on its own peak value.

regularizer has better resolution characteristics, but has slightly more noise. At locations indicated by the arrows, the proposed regularization shows noticeable improvements (better match to target).

The proposed regularization showed improved spatial resolution characteristics compared to the conventional regularization for the full-scan geometry. However, the designed impulse re-



(a) (b)

(c) (d)

Figure 3.3: Reconstructed images at end slice (a) FDK reconstruction (b) True image blurred by the target impulse response (c) Iterative reconstruction with conventional regularizer (d) Iterative reconstruction with designed regularizer

sponses do not match the target response precisely and locations with insufficient sampling still suffer from anisotropic resolution. Since 3D axial short-scans can suffer from severe anisotropy at the end slices due to their scan geometry, the proposed method may have difficulties achieving desired isotropic resolutions for short-scans.

3.2 Quadratic regularization design for isotropic and uniform noise

In addition to the regularization design method that aims for isotropic and uniform spatial resolution presented in previous section, this section introduces a method that pursues the isotropy and uniformity of noise characteristics in the reconstructed image. In practice, depending on physicians, noise characteristics may affect diagnostic quality of the reconstructed image more than the spatial resolution. Our proposed method aims to match the local power spectrum of the estimator to a target power spectrum in an efficient manner.

Consider a penalized weighted least squares (PWLS) objective function (3.1) with a quadratic regularizer defined in (2.17), where $\{r_j^l\}$ are the directional weighting coefficients that we will design.

We first define the noise covariance and correlation at each voxel. With a quadratic regularizer,

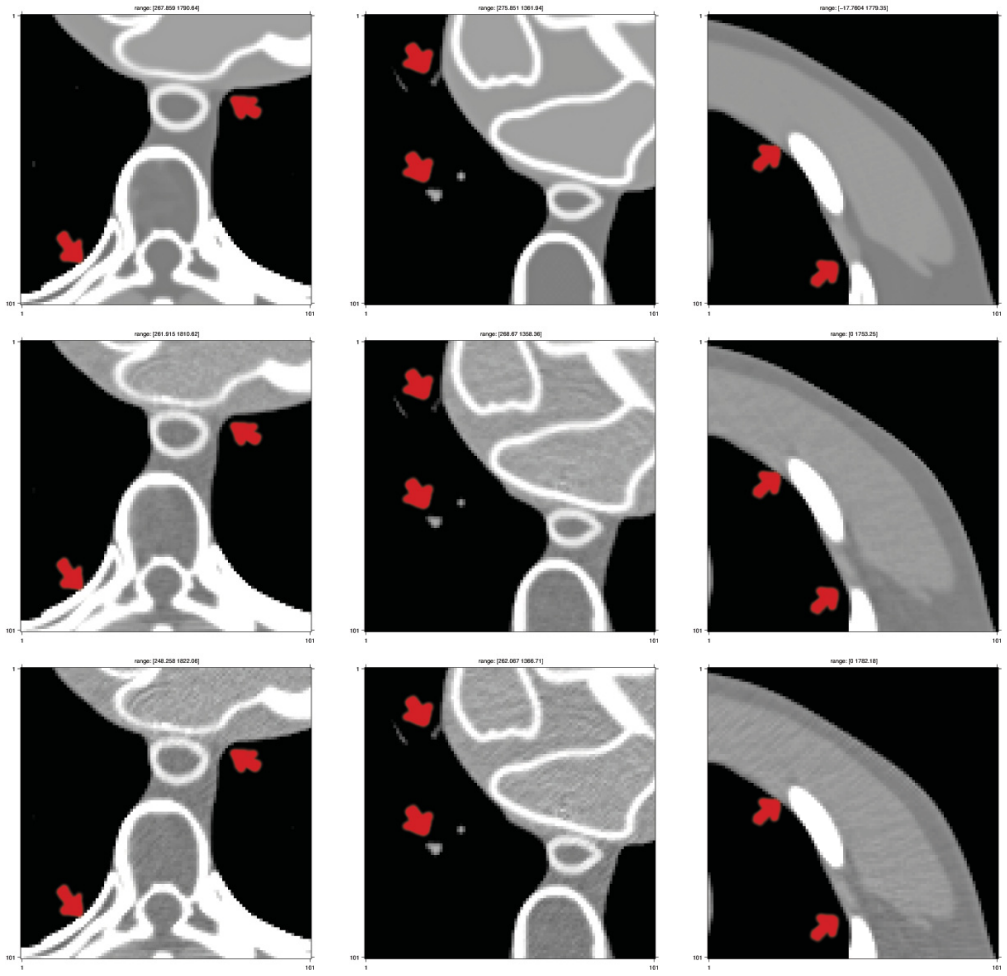


Figure 3.4: Reconstructed images with conventional regularization (middle row) and proposed regularization (bottom row) at different locations on end slices. True image blurred by the target impulse response is given as a reference (top row).

the closed-form solution of (2.8) is given by

$$\hat{\boldsymbol{x}} = \left(\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A} + \boldsymbol{R}\right)^{-1}\boldsymbol{A}'\boldsymbol{W}\boldsymbol{y}. \tag{3.33}$$

The covariance of the reconstructed image $\hat{\boldsymbol{x}}$ [39] is

$$\mathrm{cov}(\hat{\boldsymbol{x}}) = \left(\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A} + \boldsymbol{R}\right)^{-1}\boldsymbol{A}'\boldsymbol{W}\,\mathrm{cov}(\boldsymbol{y})\boldsymbol{W}\boldsymbol{A}\left(\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A} + \boldsymbol{R}\right)^{-1}. \tag{3.34}$$

If the weighting is chosen such that $\mathrm{cov}(\boldsymbol{y}) = \boldsymbol{W}^{-1}$, then the reconstructed image covariance simplifies to

$$\mathrm{cov}(\hat{\boldsymbol{x}}) = \left(\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A} + \boldsymbol{R}\right)^{-1}\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A}\left(\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A} + \boldsymbol{R}\right)^{-1}. \tag{3.35}$$

The noise property of the estimator can be expressed with the ensemble variance at each voxel:

$$\mathrm{var}\left(\hat{\boldsymbol{x}}_j\right) = \delta_j^T \left(\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A} + \boldsymbol{R}\right)^{-1} \boldsymbol{A}'\boldsymbol{W}\boldsymbol{A} \left(\boldsymbol{A}\boldsymbol{W}\boldsymbol{A} + \boldsymbol{R}\right)^{-1} \delta_j. \tag{3.36}$$

On the other hand, the width of the jth column of the covariance matrix (3.34) represents the noise correlation of jth voxel in the reconstructed image

$$\mathrm{corr}(\hat{\boldsymbol{x}}_j) = \mathrm{cov}(\hat{\boldsymbol{x}})\boldsymbol{\delta}_j = \left(\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A} + \boldsymbol{R}\right)^{-1} \boldsymbol{A}'\boldsymbol{W}\boldsymbol{A} \left(\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A} + \boldsymbol{R}\right)^{-1} \boldsymbol{\delta}_j. \tag{3.37}$$

3.2.1 Local power spectrum

We use the locally circulant approximation for $\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A}\boldsymbol{\delta}_j$ and $\boldsymbol{R}\boldsymbol{\delta}_j$ again. From (3.37), we can approximate the local noise power spectrum near voxel j as follows:

$$S^j = \frac{F\left(\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A}\boldsymbol{\delta}_j\right)}{\left(F\left(\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A}\boldsymbol{\delta}_j\right) + \beta F\left(\boldsymbol{R}\boldsymbol{\delta}_j\right)\right)^2}, \tag{3.38}$$

where $F(\cdot)$ denotes the 3-D DFT centered at jth voxel and $\boldsymbol{\delta}$ ˙j denotes a Kronecker impulse for the jth voxel.

Instead of directly using the discrete Fourier transform, we use a continuous-space analog of $H_j \triangleq F\left(\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A}\boldsymbol{\delta}_j\right)$ in spherical coordinates $\boldsymbol{\nu} \triangleq (\rho, \Phi, \Theta)$ shown in (3.10) and used for the regularization design in the previous section.

Substituting (3.10) into (3.38) yields the following expression for the continuous space analog of the noise power spectrum S^j:

$$S^j \approx \frac{M(\boldsymbol{\nu})\tilde{w}_j(\Phi)/\rho}{\left(M(\boldsymbol{\nu})\tilde{w}_j(\Phi)/\rho + \beta R^j(\rho, \Phi, \Theta)\right)^2}, \tag{3.39}$$

where $M(\boldsymbol{\nu}) = KJ(\boldsymbol{\nu})/\cos(\Theta)$ and $R^j(\rho, \Phi, \Theta)$ is the local frequency response of the regularizer near pixel j [110].

3.2.2 Target local power spectrum

Our goal is to design the regularizer such that more isotropic correlation is obtained. We approximate (3.34) as

$$\mathrm{cov}(\hat{\boldsymbol{x}}) \approx \left(\boldsymbol{D}_\eta \boldsymbol{A}'\boldsymbol{A}\boldsymbol{D}_\eta + \beta\boldsymbol{R}\right)^{-1} \boldsymbol{D}_\eta \boldsymbol{A}'\boldsymbol{A}\boldsymbol{D}_\eta \left(\boldsymbol{D}_\eta \boldsymbol{A}'\boldsymbol{A}\boldsymbol{D}_\eta + \beta\boldsymbol{R}\right)^{-1}, \tag{3.40}$$

where $\boldsymbol{D}_\eta = \mathrm{diag}\{\eta_j\}$ and $\eta_j = \sqrt{\sum_i a_{ij}^2 w_i / \sum_i a_{ij}^2}$.

With further approximations, variance at $j$the voxel becomes

$$\mathrm{var}(\hat{\boldsymbol{x}}_j) \approx \frac{1}{\eta_j^2}\left(\boldsymbol{A}'\boldsymbol{A} + \frac{\beta}{\eta_j^2}\boldsymbol{R}\right)^{-1}\boldsymbol{A}'\boldsymbol{A}\left(\boldsymbol{A}'\boldsymbol{A} + \frac{\beta}{\eta_j^2}\boldsymbol{R}\right)^{-1}. \tag{3.41}$$

From (3.41), our target, the local power spectrum at the isocenter with uniform weighting, can be written as:

$$S^0 \approx \frac{M(\boldsymbol{\nu})\tilde{u}_0(\Phi)/\rho}{\eta_0^2\left(M(\boldsymbol{\nu})\tilde{u}_0(\Phi)/\rho + (\beta/\eta_0^2)R_0(\boldsymbol{\nu})\right)^2}, \tag{3.42}$$

where $R_0(\boldsymbol{\nu})$ is the local frequency response of the regularizer at the isocenter with pre-defined directional weights, $\{r_0^l\}$, that determine the shape of the target noise power spectrum, and $\tilde{u}_0(\Phi) = |B_j(\Phi)|$.

We want to match the local noise power spectrum near jth voxel to the target local noise power spectrum, i.e., $S^j \approx S^0$.

$$S^j \approx \frac{M(\boldsymbol{\nu})\tilde{w}_j(\Phi)/\rho}{\left(M(\boldsymbol{\nu})\tilde{w}_j(\Phi)/\rho + \beta R^j(\boldsymbol{\nu})\right)^2} \approx \frac{M(\boldsymbol{\nu})\tilde{u}_0(\Phi)/\rho}{\eta_0^2\left(M(\boldsymbol{\nu})\tilde{u}_0(\Phi)/\rho + (\beta/\eta_0^2)R_0(\boldsymbol{\nu})\right)^2} \approx S^0. \tag{3.43}$$

Cross multiplying leads to the following:

$$\tilde{u}_0(\Phi)\left(M(\boldsymbol{\nu})\tilde{w}_j(\Phi)/\rho + \beta R^j(\rho,\Phi,\Theta)\right)^2 \approx \eta_0^2\tilde{w}_j(\Phi)\left(M(\boldsymbol{\nu})\tilde{u}_0(\Phi)/\rho + (\beta/\eta_0^2)R_0(\rho,\Phi,\Theta)\right)^2.$$

By taking the square root on both sides

$$\sqrt{\tilde{u}_0(\Phi)}\left(M(\boldsymbol{\nu})\tilde{w}_j(\Phi)/\rho + \beta R^j(\rho,\Phi,\Theta)\right) \approx \eta_0\sqrt{\tilde{w}_j(\Phi)}\left(M(\boldsymbol{\nu})\tilde{u}_0(\Phi)/\rho + (\beta/\eta_0^2)R^0(\rho,\Phi,\Theta)\right).$$

Finally, simplifying yields

$$R^j(\rho,\Phi,\Theta) \approx \beta^{-1}\frac{M(\boldsymbol{\nu})}{\rho}\left[\eta_0\sqrt{\tilde{u}_0(\Phi)\tilde{w}_j(\Phi)} - \tilde{w}_j(\Phi)\right] + \frac{1}{\eta_0}\sqrt{\frac{\tilde{w}_j(\Phi)}{\tilde{u}_0(\Phi)}}R_0(\boldsymbol{\nu}). \tag{3.44}$$

3.2.2.1 Regularization design for isotropic and uniform noise

Substituting (3.18) into (3.44) and simplifying yields

$$Q_j(\Phi,\Theta) \approx \beta^{-1}\frac{M(\boldsymbol{\nu})}{(2\pi\rho)^2}\left[\eta_0\sqrt{\tilde{u}_0(\Phi)\tilde{w}_j(\Phi)} - \tilde{w}_j(\Phi)\right] + \frac{1}{\eta_0}\sqrt{\frac{\tilde{w}_j(\Phi)}{\tilde{u}_0(\Phi)}}Q_0(\Phi,\Theta), \tag{3.45}$$

where $Q_j(\Phi,\Theta) \triangleq \kappa_j \sum_{l=1}^{N_l} \kappa_{l(j)} r_l^j \left(e(\Phi,\Theta)\cdot[e(\Phi_l,\Theta_l)\otimes\Delta]\right)^2$. For large ρ, the second term in (3.45) dominates:

$$Q_j(\Phi,\Theta) \approx \frac{1}{\eta_0}\sqrt{\frac{\tilde{w}_j(\Phi)}{\tilde{u}_0(\Phi)}}Q_0(\Phi,\Theta). \tag{3.46}$$

We solve the following weighted minimization problem to design the directional weighting coefficient vector $\boldsymbol{r}^j = (r_j^1,\cdots,r_j^{N_l})$ at the jth voxel

$$\boldsymbol{r}^j \triangleq \underset{\boldsymbol{r}^j\geq 0}{arg\,min} \int_0^{2\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} D_w(\Phi,\Theta)\left|\check{w}_j(\Phi,\Theta) - \sum_{l=1}^{N_l} r_j^l (e(\Phi,\Theta)\cdot[e(\Phi_l,\Theta_l)\otimes\Delta])^2\right|^2 d\Theta d\Phi, \tag{3.47}$$

where the nonnegativity constraint ensures the regularizer's convexity and we define the modified weighting function

$$\check{w}_j(\Phi,\Theta) \triangleq \frac{1}{\eta_0}\frac{\kappa_0^2}{\kappa_j^2}\sqrt{\frac{\tilde{w}_j(\Phi)}{\tilde{u}_0(\Phi)}}\sum_{l=1}^{N_l} r_0^l (e(\Phi,\Theta)\cdot[e(\Phi_l,\Theta_l)\otimes\Delta])^2, \tag{3.48}$$

where we assumed $\kappa_j \approx \kappa_l$ for l within the neighborhood of j. Notice that (3.48) is very similar to (3.24), but has a square root over the ratio $\tilde{w}_j(\Phi)/\tilde{u}_0(\Phi)$. This problem can be simplified and solved by the same procedure in previous section. We use the same choice for the weighting function, $D_w = \cos(\Theta)$.

3.2.3 Phantom experiment results

A water cylinder phantom was used to evaluate the proposed method and was scanned with a 32-slice scanner in full scan mode. The reconstruction field-of-view (FOV) was 250 mm, and the image was reconstructed by using the ICD algorithm.

Table 3.1: Comparison of image values in the reconstructed images with conventional and designed regularizers (Units: HU)

	Center Slice	End Slice
Conventional	1000 $\pm$5.22	999 $\pm$ 6.85
Designed	1000 $\pm$5.25	999 $\pm$ 5.82

Fig. 3.5 shows that the image reconstructed using the designed directional weights has slightly more uniform noise distribution, but has higher noise level. The noise level of the reconstructed image with designed directional weights is higher compared to that of conventional reconstruction because our target noise power spectrum is from uniform weighting. Selecting a proper noise power spectrum is thus important to maintain the noise level. In future works, other target noise

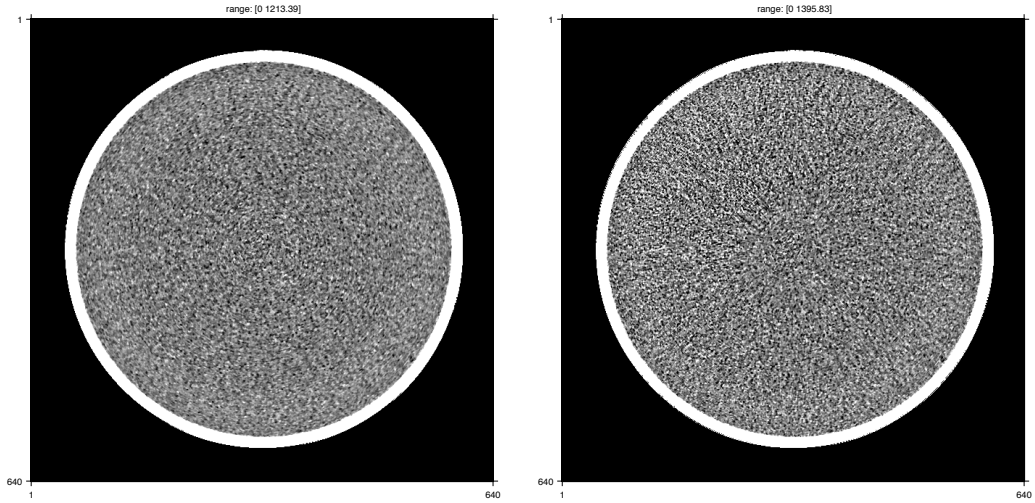


Figure 3.5: Reconstructed images at an end slice using conventional directional weights (left) and designed directional weights (right).

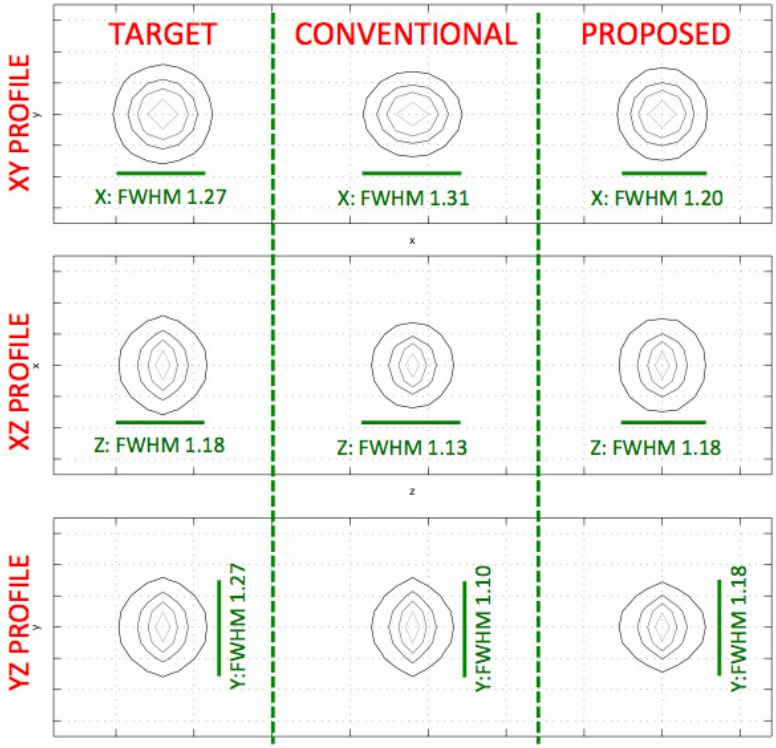


Figure 3.6: Autocorrelation of conventional regularization (middle column) and proposed regularization (right column) at (0,70,9.4) (mm). Target autocorrelation is given as a reference (left column). Each row corresponds to xy, xz, and yz profiles, respectively. Each image was normalized with its own peak value.

spectrums will be investigated to maintain the noise level and obtain further improvements. Fig. 3.6

shows that the proposed methods provide directional weights that lead to more isotropic noise characteristics in reconstructed image. Table 3.1 compares image values in the reconstructed images with conventional and designed regularizers.

Overall, the proposed method shows potential to obtain more isotropic and uniform noise characteristics for the statistical reconstruction methods. However, the proposed method suffers from the same issue as the previous design. Designed noise power spectrum does not match the target precisely, especially in the insufficiently sampled region.

3.3 Discussion

In this chapter, we proposed two regularization design methods for 3D axial CT that aim for isotropic and uniform spatial resolution or noise, respectively. Proposed methods showed promising results, and achieved reasonable isotropy and uniformity of spatial resolution or noise. However, there are many issues need to be addressed. One major problem is that the proposed methods rely on a locally circulant approximation, which fails at insufficiently sampled locations. Another issue is the limited freedom of the directional weighting coefficient design. These limitations lead to sub-optimal designs, leading to imprecise matching to the target.

Another limitation of these methods is that we only focused on quadratic regularization. Edge-preserving regularization is the preferred method for denoising in CT problems, so it would be desirable to extend the proposed method to edge-preserving regularization. How to use the designed directional weights in edge-preserving regularization needs to be investigated.

How to choose the target is still an open question, especially for the noise power spectrum. Matching to one fixed target may be a too strong constraint considering the issues mentioned above. Combining these two methods to control both spatial resolution and noise characteristics may be a possible future work. The approximation for H_j [110] used in this chapter may require further refinements to estimate H_j more accurately. As a future work, we will use this refined approximation to obtain better isotropy and uniformity of spatial resolution and noise characteristics of the estimator. Furthermore, the minimization (3.29) should consider various constraints to improve designed solutions.

CHAPTER IV

Regularization designs using the hypothetical geometry

As mentioned previously, conventional regularized image reconstruction leads to spatially variant spatial resolution and noise characteristics because of interactions between the system models and the regularization. Previous regularization design methods aiming to solve such issues mostly rely on circulant approximations of the Fisher information matrix that are very inaccurate for undersampled geometries like short-scan cone-beam CT. This chapter extends the regularization method proposed in [42] to 3D cone-beam CT by introducing a hypothetical scanning geometry that helps address the sampling properties. The proposed regularization designs were compared with the original method in [42] with both phantom simulation and clinical reconstruction in 3D axial X-ray CT. The proposed regularization methods yield improved spatial resolution or noise uniformity in statistical image reconstruction for short-scan axial cone-beam CT.

In [42], a regularizer based on the aggregated certainty was developed for 2D PET to yield images with approximately uniform spatial resolution, and that regularizer has been used for other geometries and modalities [84, 90, 98, 99, 105]. However, the aggregated certainty regularizer does not provide uniform resolution when applied to modalities such as 2D short-scan fan-beam CT or 3D cone beam CT because of asymmetries in scan geometry, such as cone angle and short-scan orbit. In [99] and [84], the original aggregated certainty regularizer was modified with a diagonal scaling factor for 3D PET. Recently, it was also extended to both static and multi-frame reconstruction in 3D PET by considering spatially variant and frame-dependent sensitivity [66]. Since the term aggregated certainty maybe misleading for some imaging modalities such as CT, instead, we use more general term "pre-tuned spatial strength", which represents that the purpose of the function is to control the regularization strength at each voxel, before the reconstruction process, so that the reconstructed image is guided to have desired characteristics, *i.e.*, uniform resolution.

Many previous regularization design methods focussed on choosing directional coefficients in

This chapter is based on [21, 22]

the regularizer by matching local characteristics, such as impulse response or correlation function, of the estimator to target characteristics to achieve uniform and isotropic resolution [19, 116, 118] or noise characteristics [20, 108]. Since both the regularization parameter and the pre-tuned spatial strength can be incorporated into directional coefficients, those regularization design methods are more general and flexible than simply adjusting the regularization strength at each voxel. However, such design methods require additional computations to design the coefficients for every voxel, and it is challenging to obtain both uniformity and isotropy at the same time for either spatial resolution or noise characteristics. Especially for the undersampled voxels in cone-beam CT, locally circulant approximations of the Fisher information matrix are very inaccurate, leading to imperfect coefficient designs at such locations. Furthermore, the memory requirement to store all directional weights for every voxel can be burdensome.

The goal of this chapter is to refine the regularizer $\mathsf{R}(\boldsymbol{x})$, by designing $\{\kappa_j\}$ so that the reconstructed image has more uniform resolution or noise properties. We want to improve upon the design proposed in [42] which was

$$\kappa_j = \sqrt{\frac{\sum_{i=1}^{n_\mathrm{d}} a_{ij}^2 w_i}{\sum_{i=1}^{n_\mathrm{d}} a_{ij}^2}}. \tag{4.1}$$

We continue to refer (4.1) as the (original) aggregated certainty to distinguish from our newly designed pre-tuned spatial strengths.

4.1 Metrics to analyze spatial resolution and noise properties

In this section, we review the metrics to analyze spatial resolution and noise properties in the reconstructed image.

One common metric for measuring the local resolution is the width of the local impulse response at jth voxel, such as the full width half maximum (FWHM) [42]. As an alternative, the peak amplitude of the local impulse response, or the contrast recovery coefficient (CRC) [98], can be used to quantify resolution

$$\mathrm{CRC}(\boldsymbol{l}_j) \triangleq \boldsymbol{\delta}_j^T \boldsymbol{l}_j = \boldsymbol{\delta}_j^T [\boldsymbol{A}' \boldsymbol{W} \boldsymbol{A} + \boldsymbol{R}]^{-1} \boldsymbol{A}' \boldsymbol{W} \boldsymbol{A} \boldsymbol{\delta}_j. \tag{4.2}$$

To measure isotropy of the impulse response, the width measure will be more effective. On the other hand, uniformity of the impulse response is easier to assess with the CRC. In this chapter, we use the CRC of the local impulse response to quantify spatial resolution.

When the statistical weighting is chosen such that $\mathrm{cov}(\boldsymbol{y}) = \boldsymbol{W}^{-1}$, the reconstruction image covariance is given by (3.34). However, in some cases, additional factors are applied to $\boldsymbol{W}$. For

instance, Parker weighting [95] is applied to sinogram measurements for short-scan FBP reconstruction. It may also be used in iterative reconstructions so that the temporal resolution of such reconstructions would match that of FBP reconstruction. Additional weightings change the statistical characteristics of the weighting $\boldsymbol{W}$, thus it no longer satisfies $\mathrm{cov}(\boldsymbol{y}) = \boldsymbol{W}^{-1}$. We can express the statistical weighting more generally as

$$\boldsymbol{W} = \hat{\boldsymbol{W}}\boldsymbol{V} = \boldsymbol{D}[\hat{w}_i v_i], \tag{4.3}$$

where the weighting $\hat{\boldsymbol{W}}$ is the conventional choice that satisfies $\hat{w}_i = 1/\mathsf{Var}\{y_i\}$, and $\{v_i\}$ are the additional weighting elements. The reconstructed image covariance (3.34) with the new statistical weighting can be expressed as follows

$$\mathrm{cov}(\hat{\boldsymbol{x}}) = \left(\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A} + \boldsymbol{R}\right)^{-1} \boldsymbol{A}'\check{\boldsymbol{W}}\boldsymbol{A} \left(\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A} + \boldsymbol{R}\right)^{-1}, \tag{4.4}$$

where $\check{\boldsymbol{W}} = \boldsymbol{V}\hat{\boldsymbol{W}}\boldsymbol{V} = \boldsymbol{D}[\hat{w}_i v_i^2]$.

The noise property of the estimator can be quantified with the ensemble variance at each voxel:

$$\mathrm{var}\left(\hat{\boldsymbol{x}}_j\right) = \delta_j^T \left(\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A} + \boldsymbol{R}\right)^{-1} \boldsymbol{A}'\check{\boldsymbol{W}}\boldsymbol{A}' \left(\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A} + \boldsymbol{R}\right)^{-1} \boldsymbol{\delta}_j. \tag{4.5}$$

The entire jth column of the covariance matrix (4.4) represents the noise correlation of the jth voxel in the reconstructed image with all other voxels

$$\mathrm{corr}(\hat{\boldsymbol{x}}_j) \triangleq \left(\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A} + \boldsymbol{R}\right)^{-1} \boldsymbol{A}'\check{\boldsymbol{W}}\boldsymbol{A} \left(\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A} + \boldsymbol{R}\right)^{-1} \boldsymbol{\delta}_j. \tag{4.6}$$

4.2 New regularization designs using the hypothetical geometry

In this section, we review the aggregated certainty developed in [42], and then modify it to develop new regularization designs that provide approximately uniform resolution or noise properties by using an "ideal" system matrix factorization.

4.2.1 System matrix generalization using the hypothetical geometry

The aggregated certainty in [42] was developed for shift-invariant systems like 2D PET. For shift-variant systems like CBCT, the formulation in [42] must be modified.

The Fisher information matrix $\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A}$ is shift-variant for both emission and transmission tomography, causing nonuniform properties of the reconstructed image. Statistical weighting is only partially responsible for the nonuniformity; the Fisher information matrix for unweighted case $\boldsymbol{A}'\boldsymbol{A}$ is also shift-variant for 3D PET and CT. In [42], the system matrix $\boldsymbol{A}$ was factored into three

elements as follows

$$a_{ij} = c_i g_{ij} s_j, \tag{4.7}$$

where $\{c_i\}$ denote ray-dependent factors, $\{s_j\}$ denote voxel-dependent factors, and $\boldsymbol{G} = [g_{ij}]$ represents the object-independent geometric portion of the tomographic system response. Ideally we would like to choose $\{c_i\}$ and $\{s_j\}$ and $\boldsymbol{G}$ such that $\boldsymbol{G}'\boldsymbol{G}$ is shift-invariant. Shift-invariant $\boldsymbol{G}'\boldsymbol{G}$ can be more accurately approximated with a circulant matrix (implemented via fast Fourier transform (FFT)), which will lead to improved regularizer designs. The matrix representation of (4.7) is

$$\boldsymbol{A} = \boldsymbol{D}[c_i]\boldsymbol{G}\boldsymbol{D}[s_j]. \tag{4.8}$$

Since this representation is not unique, we can design each of the factors to make $\boldsymbol{G}'\boldsymbol{G}$ "very shift-invariant". The original design presented in [42] for PET assumed uniform voxel-dependent factors, *i.e.*, $s_j = 1, \forall j$, and the ray-dependent part included only non-geometric aspects such as detector efficiency and dead time. In 2D PET, this leads to geometric factors $\{g_{ij}\}$ for which $\boldsymbol{G}'\boldsymbol{G}$ is nearly shift-invariant. However, this conventional choice of $\boldsymbol{G}$ leads to $\boldsymbol{G}'\boldsymbol{G}$ that is highly shift-variant for 3D cone beam CT and even for 2D fan-beam CT when short-scan geometry is considered.

Here, we present a new generalization of the form (4.7) that works in various geometries including 3D cone beam CT. First, we consider the geometric sampling properties of $\boldsymbol{A}$ and consider what rays are "missing" that cause $\boldsymbol{A}'\boldsymbol{A}$ to be shift-variant. For example, in a short-scan fan-beam geometry we are missing some of the views that would have been acquired with a full 360° scan. As another example, for a CBCT axial scan, we are missing the data that would have been acquired with a "step and shoot" set of axial scans [56]. For axial CBCT with a full 360° scan, $\boldsymbol{A}'\boldsymbol{A}$ is approximately shift-invariant over the fully-sampled, so-called "football region", it is natural to define $\boldsymbol{G}$ to be a hypothetical system matrix having extra detector rows such that the entire reconstruction volume is contained in its corresponding football region[4]. Fig. 4.1 illustrates some cases described above. In general, we define $\boldsymbol{G}$ to be some $N_\mathrm{g} \times n_\mathrm{p}$ system matrix corresponding to an ideal, fully sampled geometry, for which $\boldsymbol{G}'\boldsymbol{G}$ is approximately shift-invariant. The matrix $\boldsymbol{G}$ has the same number of columns as $\boldsymbol{A}$ but has more rows ($N_\mathrm{g} > n_\mathrm{d}$); the rows of $\boldsymbol{A}$ are a subset of the rows of $\boldsymbol{G}$.

Second, we replace the usual diagonal matrix $\boldsymbol{D}[c_i]$ in (4.8) with $\boldsymbol{D}[c_i]\boldsymbol{P}$ where $\boldsymbol{P}$ is a $n_\mathrm{d} \times N_\mathrm{g}$ matrix that that selects the rows of the hypothetical geometry $\boldsymbol{G}$ corresponding to those of the actual geometry $\boldsymbol{A}$. Each row of $\boldsymbol{P}$ is entirely zero except for a single element that is unity. By ordering the rows of $\boldsymbol{G}$ appropriately, we can use $\boldsymbol{P} = [\boldsymbol{I}_{n_\mathrm{d}} \; \boldsymbol{0}_{n_\mathrm{d} \times (N_\mathrm{g} - n_\mathrm{d})}]$. An important property

[4] A hypothetical parallel-beam geometry is another option for $\boldsymbol{G}$ that may lead to $\boldsymbol{G}'\boldsymbol{G}$ that is even more shift-invariant; that choice would require an additional cone-to-parallel rebinning process [2].

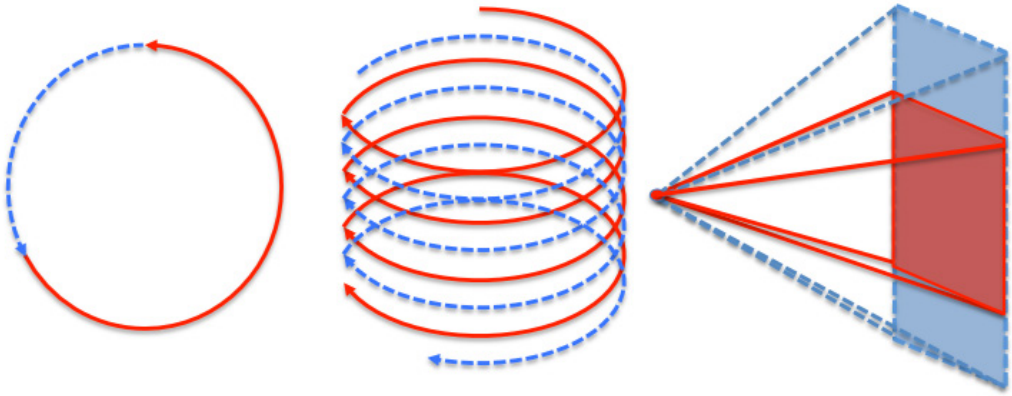


Figure 4.1: Illustration of possible hypothetical geometries $\boldsymbol{G}$. Short-scan orbit vs. full-scan orbit (left). Intertwined multiple helical geometries (middle). Cone-beam with extended detector rows (right).

of the row selection matrix $\boldsymbol{P}$ is that $\boldsymbol{P}'\boldsymbol{W}\boldsymbol{P}$ is a $N_{\mathrm{g}} \times N_{\mathrm{g}}$ diagonal matrix where each element corresponds to a w_i value for actual rays and is zero for the hypothetical rays.

With this generalization, we can rewrite the Fisher information matrix of the data fitting term as follows:

$$\begin{aligned} \boldsymbol{A}'\boldsymbol{W}\boldsymbol{A} &= \boldsymbol{D}[s_j]\boldsymbol{G}'\boldsymbol{P}'\boldsymbol{D}[c_i]\boldsymbol{W}\boldsymbol{D}[c_i]\boldsymbol{P}\boldsymbol{G}\boldsymbol{D}[s_j] \\ &= \boldsymbol{D}[s_j]\boldsymbol{G}'\boldsymbol{P}'\boldsymbol{D}[w_i c_i^2]\boldsymbol{P}\boldsymbol{G}\boldsymbol{D}[s_j]. \end{aligned} \tag{4.9}$$

Since the Fisher information matrix is fairly concentrated near its diagonal elements [42], we approximate (4.9) as

$$\begin{aligned} \boldsymbol{A}'\boldsymbol{W}\boldsymbol{A} &\approx \boldsymbol{D}[s_j]\boldsymbol{D}[\eta_j]\boldsymbol{G}'\boldsymbol{G}\boldsymbol{D}[\eta_j]\boldsymbol{D}[s_j] \\ &= \boldsymbol{D}[\lambda_j]\boldsymbol{G}'\boldsymbol{G}\boldsymbol{D}[\lambda_j] = \Lambda\boldsymbol{G}'\boldsymbol{G}\Lambda, \end{aligned} \tag{4.10}$$

where the following factors match the diagonals of (4.10):

$$\eta_j = \sqrt{\frac{\sum_{i=1}^{n_{\mathrm{d}}} g_{ij}^2 c_i^2 w_i}{\sum_{i=1}^{N_{\mathrm{g}}} g_{ij}^2}}, \tag{4.11}$$

$$\lambda_j = s_j\eta_j = \sqrt{\frac{\sum_{i=1}^{n_{\mathrm{d}}} a_{ij}^2 w_i}{\sum_{i=1}^{N_{\mathrm{g}}} g_{ij}^2}}, \tag{4.12}$$

$$\Lambda \triangleq \boldsymbol{D}[\lambda_j]. \tag{4.13}$$

Different choices of c_i and s_j will lead to various designs for $\boldsymbol{G} = \{g_{ij}\}$. As an example, for axial cone-beam CT, assuming $c_i = 1, \forall i$, and $s_j = 1, \forall j$, will lead to $\boldsymbol{G}$ that corresponds to another axial cone-beam CT geometry with extended detector rows and full 360 degree orbit. For some other choices, we may not have a physical interpretation for the system represented by $\boldsymbol{G}$.

Both ray- and voxel-dependent factors need to be designed based on different criteria depending on the modality, *i.e.*, for SPECT, s_j should be designed to properly model the nonuniform spatial sensitivity and, for PET, c_i to represent detector characteristics.

4.2.2 Regularization with uniform resolution property

Substituting (4.10) into (3.7) yields the following approximation for the local impulse response at jth voxel:

$$\begin{aligned} \boldsymbol{l}_j &\approx \left(\Lambda \boldsymbol{G}'\boldsymbol{G}\Lambda + \boldsymbol{R}\right)^{-1} \Lambda \boldsymbol{G}'\boldsymbol{G}\Lambda \boldsymbol{\delta}_j \\ &= \Lambda^{-1} \left(\boldsymbol{G}'\boldsymbol{G} + \Lambda^{-1}\boldsymbol{R}\Lambda^{-1}\right)^{-1} \boldsymbol{G}'\boldsymbol{G}\Lambda \boldsymbol{\delta}_j. \end{aligned} \tag{4.14}$$

Typically, the local impulse response $\boldsymbol{l}_j$ is concentrated about voxel j and clearly $\Lambda\boldsymbol{\delta}_j = \lambda_j \boldsymbol{\delta}_j$. So we approximate (4.14) as the following final expression for the local impulse response:

$$\boldsymbol{l}_j \approx \left(\boldsymbol{G}'\boldsymbol{G} + \lambda_j^{-2}\boldsymbol{R}\right)^{-1} \boldsymbol{G}'\boldsymbol{G}\boldsymbol{\delta}_j. \tag{4.15}$$

Now that we have analyzed the local impulse response, we focus on designing the coefficients $\{\kappa_j\}$ in the regularizer (2.17); these coefficients affect the Hessian $\boldsymbol{R}$ in (4.15) and thus control the spatial resolution. Our goal here is to choose $\{\kappa_j\}$ to provide approximately uniform spatial resolution by matching the local impulse response at jth voxel, $\boldsymbol{l}_j$, to a target local impulse response, $\boldsymbol{l}_{\mathrm{ref}}$, *i.e.*, $\boldsymbol{l}_j \approx \boldsymbol{l}_{\mathrm{ref}}$. Using the approximation (4.15), we write the target local impulse response at a reference point, such as the isocenter, as follows:

$$\boldsymbol{l}_{\mathrm{ref}} \triangleq \left(\boldsymbol{G}'\boldsymbol{G} + \boldsymbol{R}_0\right)^{-1} \boldsymbol{G}'\boldsymbol{G}\boldsymbol{\delta}_{\mathrm{ref}}, \tag{4.16}$$

where $\boldsymbol{R}_0$ is the Hessian of a regularizer $\mathsf{R}_0(\boldsymbol{x})$ that provides desirable spatial resolution property at the reference point. $\mathsf{R}_0(\boldsymbol{x})$ has the same form as (2.17) but possibly with a different set of $\{\kappa_j\}$ or $\kappa_j = 1, \forall j$.

Our design for $\boldsymbol{G}$ leads to locally circulant $\boldsymbol{G}'\boldsymbol{G}\boldsymbol{\delta}_j$, and we assume $\boldsymbol{R}\boldsymbol{\delta}_j$ is approximately locally shift-invariant [99]. Taking Fourier transform of (4.15) yields the following expression for

the local frequency response

$$L^j \approx \frac{F\left(\boldsymbol{G}'\boldsymbol{G}\boldsymbol{\delta}_j\right)}{F\left(\boldsymbol{G}'\boldsymbol{G}\boldsymbol{\delta}_j\right) + \lambda_j^{-2} F\left(\boldsymbol{R}\boldsymbol{\delta}_j\right)}, \tag{4.17}$$

where $F(\cdot)$ denotes the 3-D DFT centered at jth voxel. We match the local frequency response of jth voxel to the target frequency response, i.e.,

$$\begin{aligned} L^j &\approx \frac{F\left(\boldsymbol{G}'\boldsymbol{G}\boldsymbol{\delta}_j\right)}{F\left(\boldsymbol{G}'\boldsymbol{G}\boldsymbol{\delta}_j\right) + \lambda_j^{-2} F\left(\boldsymbol{R}\boldsymbol{\delta}_j\right)} \\ &\approx \frac{F\left(\boldsymbol{G}'\boldsymbol{G}\boldsymbol{\delta}_{\mathrm{ref}}\right)}{F\left(\boldsymbol{G}'\boldsymbol{G}\boldsymbol{\delta}_{\mathrm{ref}}\right) + F\left(\boldsymbol{R}_0\boldsymbol{\delta}_{\mathrm{ref}}\right)} \approx L^{\mathrm{ref}}. \end{aligned} \tag{4.18}$$

Cross multiplying and simplifying yields

$$F\left(\boldsymbol{R}\boldsymbol{\delta}_j\right) F\left(\boldsymbol{G}'\boldsymbol{G}\boldsymbol{\delta}_{\mathrm{ref}}\right) \approx \lambda_j^2 F\left(\boldsymbol{R}_0\boldsymbol{\delta}_{\mathrm{ref}}\right) F\left(\boldsymbol{G}'\boldsymbol{G}\boldsymbol{\delta}_j\right). \tag{4.19}$$

We design $\{\kappa_j\}$ by minimizing the least squares difference between both sides of (4.19)

$$\hat{\kappa}_j = \arg\min_{\kappa_j} \| F\left(\boldsymbol{R}\boldsymbol{\delta}_j\right) F\left(\boldsymbol{G}'\boldsymbol{G}\boldsymbol{\delta}_{\mathrm{ref}}\right) - \lambda_j^2 F\left(\boldsymbol{R}_0\boldsymbol{\delta}_{\mathrm{ref}}\right) F\left(\boldsymbol{G}'\boldsymbol{G}\boldsymbol{\delta}_j\right) \|^2. \tag{4.20}$$

For the quadratic potential function, the local frequency response of the regularizer Hessian is [116]

$$F\left(\boldsymbol{R}\boldsymbol{\delta}_j\right) = \kappa_j^2 R_\omega, \tag{4.21}$$

$$R_\omega \triangleq \sum_{l=1}^{N_l} \beta r^l \left| C_l(\omega_1, \omega_2, \omega_3) \right|^2, \tag{4.22}$$

using the usual approximation $\kappa_j \approx \kappa_l$ for l within the neighborhood of j, where C_l denotes the discrete-space Fourier transform of $c_l[n, m, z]$. Without loss of generality, we can choose $\mathsf{R}_0(\boldsymbol{x})$ such that

$$F\left(\boldsymbol{R}_0\boldsymbol{\delta}_{\mathrm{ref}}\right) = \sum_{l=1}^{N_l} \beta r^l \left| C_l(\omega_1, \omega_2, \omega_3) \right|^2 = R_\omega. \tag{4.23}$$

Substituting (4.21) and (4.23) into (4.20) yields the following simplified expression:

$$\hat{\kappa}_j = \arg\min_{\kappa_j} \| \kappa_j^2 R_\omega F\left(\boldsymbol{G}'\boldsymbol{G}\boldsymbol{\delta}_{\mathrm{ref}}\right) - \lambda_j^2 R_\omega F\left(\boldsymbol{G}'\boldsymbol{G}\boldsymbol{\delta}_j\right) \|^2. \tag{4.24}$$

Solving (4.24), we obtain

$$\hat{\kappa}_j = \lambda_j \frac{\sqrt{\langle R_\omega F\left(\boldsymbol{G}'\boldsymbol{G}\boldsymbol{\delta}_{\text{ref}}\right), R_\omega F\left(\boldsymbol{G}'\boldsymbol{G}\boldsymbol{\delta}_j\right)\rangle}}{\|R_\omega F\left(\boldsymbol{G}'\boldsymbol{G}\boldsymbol{\delta}_{\text{ref}}\right)\|}, \tag{4.25}$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product for 3D DFT space.

When $\boldsymbol{G}'\boldsymbol{G}$ is approximately shift-invariant, we have

$$F\left(\boldsymbol{G}'\boldsymbol{G}\boldsymbol{\delta}_j\right) \approx F\left(\boldsymbol{G}'\boldsymbol{G}\boldsymbol{\delta}_{\text{ref}}\right), \tag{4.26}$$

and (4.25) simplifies to $\kappa_j = \lambda_j$.

The presented design process can address more general purposes besides obtaining resolution uniformity. For instance, one may want to match a spatially varying target response that depends on certain characteristics, such as the sampling at each voxel, so that each voxel would have different resolution properties for specific purposes.

Our new regularizer for uniform resolution properties in the reconstructed image (hereafter R-REG) is given by (2.17) with

$$\hat{\kappa}_j = \sqrt{\frac{\sum_{i=1}^{n_{\text{d}}} a_{ij}^2 w_i}{\sum_{i=1}^{N_{\text{g}}} g_{ij}^2}}. \tag{4.27}$$

This new pre-tuned spatial strength function (4.27) has a very similar form to that of the original certainty (4.1) proposed in [42], but with a different denominator. This new denominator takes effect when the voxel j is at an undersampled location. When it is fully-sampled, the new pre-tuned spatial strength is exactly the same as the original certainty since $\sum_{i=1}^{N_{\text{g}}} g_{ij}^2 = \sum_{i=1}^{n_{\text{d}}} a_{ij}^2$ for such locations. For undersampled region, this new denominator decreases the regularization strength, leading to sharper and possibly noisier reconstructed images compared to using the original aggregated certainty (4.1). For practical purpose, we can approximate (4.27) as follows

$$\hat{\kappa}_j \approx \sqrt{\frac{\sum_{i=1}^{n_{\text{d}}} a_{ij} w_i}{\sum_{i=1}^{N_{\text{g}}} g_{ij}}}. \tag{4.28}$$

Unlike the back-projection of the statistical weighting, $\sum_{i=1}^{n_{\text{d}}} a_{ij} w_i$, calculating the sum of rows of the Hessian $\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A}$, $\sum_{i=1}^{n_{\text{d}}} a_{ij}^2 w_i$, is sometimes not available or easily implementable. We can obtain reasonably similar results using the approximation (see Appendix A for empirical verification).

Even though the new regularizer design was derived for quadratic regularization, it can be also applied to regularizations with non-quadratic potential functions, following the spirit of [7]. For such regularizations, the Hessian of the regularization depends on $\boldsymbol{x}$, i.e., $\boldsymbol{R}(\boldsymbol{x}) = \nabla^2 \mathsf{R}(\boldsymbol{x})$, and

the approximate local impulse response can be written as

$$\boldsymbol{l}_j \approx \left(\boldsymbol{G}'\boldsymbol{G} + \frac{1}{\lambda_j^2}\boldsymbol{R}(\boldsymbol{x}) \right)^{-1} \boldsymbol{G}'\boldsymbol{G}\boldsymbol{\delta}_j. \tag{4.29}$$

Since $\boldsymbol{D}[\hat{\kappa_j}] = \Lambda$, if we assume that $\hat{\kappa}_j$ is almost the same within its neighborhood, the Hessian of the regularization (2.17) can be approximated as follows [42] The Hessian of the regularization (2.17) is given by [42]

$$\boldsymbol{R}_{jk}(\boldsymbol{x}) = \left\{ \begin{array}{ll} 2\sum_{j \neq l} \hat{\kappa}_l \hat{\kappa}_j \beta_j^l \ddot{\psi}_l(\boldsymbol{x}_j - \boldsymbol{x}_l), & j = l \\ -2\hat{\kappa}_l \hat{\kappa}_j \beta_j^l \ddot{\psi}_l(\boldsymbol{x}_j - \boldsymbol{x}_l), & j \neq l \end{array} \right\}, \tag{4.30}$$

and since $\boldsymbol{D}[\hat{\kappa_j}] = \Lambda$ we can approximate $\boldsymbol{R}(\boldsymbol{x})$ as follows

$$\boldsymbol{R}(\boldsymbol{x}) \approx \Lambda \boldsymbol{R}^*(\boldsymbol{x}) \Lambda, \tag{4.31}$$

where $\boldsymbol{R}^*(\boldsymbol{x})$ is the Hessian of the following regularization without the pre-tuned spatial strength

$$\mathsf{R}^*(\boldsymbol{x}) = \sum_{j=1}^{n_{\mathrm{p}}} \sum_{l=1}^{N_l} \beta r^l \psi_l \left((c_l * * * x)\,[n, m, z] \right). \tag{4.32}$$

Substituting (4.31) to (4.14) leads the following new approximation for the local impulse response

$$\boldsymbol{l}_j \approx \left(\boldsymbol{G}'\boldsymbol{G} + \boldsymbol{R}^*(\boldsymbol{x}^{\mathrm{true}}) \right)^{-1} \boldsymbol{G}'\boldsymbol{G}\boldsymbol{\delta}_j. \tag{4.33}$$

Since $\boldsymbol{G}'\boldsymbol{G}$ was designed to be closely shift-invariant, the proposed regularization provides approximately uniform resolution at every voxel. However, due to the object-dependence of the Hessian of the regularization $\boldsymbol{R}^*(\boldsymbol{x}^{\mathrm{true}})$, the proposed regularization leads to some non-uniformities in the reconstructed image such as less smoothing on the edges, *i.e.*, edge-preserving regularization. This non-uniformity is an important characteristic of the non-quadratic regularization, and it will be preserved.

The proposed regularizer (4.27) attempts to address non-uniformities caused by both the shift-variant scanning geometries and by interactions between the regularization and the statistical weights. The derivation assumed that κ_j changes very slowly within its neighborhood. However, this assumption may fail for certain regions such as near the edges of a structure. Furthermore, since we are only adjusting the "overall strength" of the regularization at each voxel and not its "directional strength" for each neighboring voxel, the proposed regularization cannot correct for asymmetry in local impulse responses. The proposed regularization is designed to generate uniform spatial resolution in terms of CRC. To obtain isotropic local impulse response, one would

need to design the directional coefficients, $\{r_j^l\}$, at each location [19].

4.2.3 Regularization with uniform noise property

Using the approximation (4.10), we approximate the local noise correlation $\mathrm{corr}(\hat{\boldsymbol{x}})$ as follows

$$\mathrm{corr}(\hat{\boldsymbol{x}}_j) \approx \left(\bar{\Lambda}\boldsymbol{G}'\boldsymbol{G}\bar{\Lambda} + \boldsymbol{R}\right)^{-1} \hat{\Lambda}\boldsymbol{G}'\boldsymbol{G}\hat{\Lambda} \left(\bar{\Lambda}\boldsymbol{G}'\boldsymbol{G}\bar{\Lambda} + \boldsymbol{R}\right)^{-1} \boldsymbol{\delta}_j, \tag{4.34}$$

where

$$\bar{\lambda}_j \triangleq \sqrt{\frac{\sum_{i=1}^{n_\mathrm{d}} a_{ij}^2 \hat{w}_i v_i}{\sum_{i=1}^{N_\mathrm{g}} g_{ij}^2}}, \quad \bar{\Lambda} \triangleq \boldsymbol{D}[\bar{\lambda}_j], \tag{4.35}$$

$$\hat{\lambda}_j \triangleq \sqrt{\frac{\sum_{i=1}^{n_\mathrm{d}} a_{ij}^2 \hat{w}_i v_i^2}{\sum_{i=1}^{N_\mathrm{g}} g_{ij}^2}}, \quad \hat{\Lambda} \triangleq \boldsymbol{D}[\hat{\lambda}_j], \tag{4.36}$$

and v_i and $\hat{w}_i$ were introduced in (4.3).

We further approximate (4.34) as

$$\begin{aligned} \mathrm{corr}(\hat{\boldsymbol{x}}_j) &\approx \bar{\Lambda}^{-1} \left(\boldsymbol{G}'\boldsymbol{G} + \bar{\Lambda}^{-1}\boldsymbol{R}\bar{\Lambda}^{-1}\right)^{-1} \bar{\Lambda}^{-1}\hat{\Lambda}\boldsymbol{G}'\boldsymbol{G}\hat{\Lambda}\bar{\Lambda}^{-1} \left(\boldsymbol{G}'\boldsymbol{G} + \bar{\Lambda}^{-1}\boldsymbol{R}\bar{\Lambda}^{-1}\right)^{-1} \bar{\Lambda}^{-1}\boldsymbol{\delta}_j \\ &\approx \frac{\hat{\lambda}_j^2}{\bar{\lambda}_j^4} \left(\boldsymbol{G}'\boldsymbol{G} + \bar{\lambda}_j^{-2}\boldsymbol{R}\right)^{-1} \boldsymbol{G}'\boldsymbol{G} \left(\boldsymbol{G}'\boldsymbol{G} + \bar{\lambda}_j^{-2}\boldsymbol{R}\right)^{-1} \boldsymbol{\delta}_j, \end{aligned} \tag{4.37}$$

using the usual assumption that the local noise correlation $\mathrm{corr}(\hat{\boldsymbol{x}})$ is concentrated about voxel j. From (4.37), the local noise power spectrum of voxel j is approximately

$$S^j \approx \frac{\hat{\lambda}_j^2}{\bar{\lambda}_j^4} \frac{F\left(\boldsymbol{G}'\boldsymbol{G}\boldsymbol{\delta}_j\right)}{\left(F\left(\boldsymbol{G}'\boldsymbol{G}\boldsymbol{\delta}_j\right) + \bar{\lambda}_j^{-2} F\left(\boldsymbol{R}\boldsymbol{\delta}_j\right)\right)^2}. \tag{4.38}$$

To obtain uniform noise properties, we match the local noise correlation at the jth voxel to a target noise correlation in the frequency domain, i.e., $S^j \approx S^{\mathrm{ref}}$. For the target, we use the local noise correlation at the reference point and assume that the regularizer $\mathsf{R}_0(\boldsymbol{x})$ was chosen to provide the proper noise level at that location. Our design goal becomes:

$$\frac{\hat{\lambda}_j^2}{\bar{\lambda}_j^4} \frac{F\left(\boldsymbol{G}'\boldsymbol{G}\boldsymbol{\delta}_j\right)}{\left(F\left(\boldsymbol{G}'\boldsymbol{G}\boldsymbol{\delta}_j\right) + \bar{\lambda}_j^{-2} F\left(\boldsymbol{R}\boldsymbol{\delta}_j\right)\right)^2} \approx \frac{\hat{\lambda}_{\mathrm{ref}}^2}{\bar{\lambda}_{\mathrm{ref}}^4} \frac{F\left(\boldsymbol{G}'\boldsymbol{G}\boldsymbol{\delta}_{\mathrm{ref}}\right)}{\left(F\left(\boldsymbol{G}'\boldsymbol{G}\boldsymbol{\delta}_{\mathrm{ref}}\right) + F\left(\boldsymbol{R}_0\boldsymbol{\delta}_{\mathrm{ref}}\right)\right)^2}. \tag{4.39}$$

Cross multiplying leads to

$$\hat{\lambda}_j^2 \bar{\lambda}_{\text{ref}}^4 F\left(\boldsymbol{G}'\boldsymbol{G}\boldsymbol{\delta}_j\right)\left(F\left(\boldsymbol{G}'\boldsymbol{G}\boldsymbol{\delta}_{\text{ref}}\right)+F\left(\boldsymbol{R}_0\boldsymbol{\delta}_{\text{ref}}\right)\right)^2 \\ \approx \hat{\lambda}_{\text{ref}}^2 \bar{\lambda}_j^4 F\left(\boldsymbol{G}'\boldsymbol{G}\boldsymbol{\delta}_{\text{ref}}\right)\left(F\left(\boldsymbol{G}'\boldsymbol{G}\boldsymbol{\delta}_j\right)+\bar{\lambda}_j^{-2} F\left(\boldsymbol{R}\boldsymbol{\delta}_j\right)\right)^2. \quad (4.40)$$

Using (4.21) and (4.23), we simplify (4.40) to

$$\kappa_j^2 \sqrt{F\left(\boldsymbol{G}'\boldsymbol{G}\boldsymbol{\delta}_{\text{ref}}\right)} R_\omega \\ \approx \left(\frac{\hat{\lambda}_j \bar{\lambda}_{\text{ref}}^2}{\hat{\lambda}_{\text{ref}}}\right)\sqrt{F\left(\boldsymbol{G}'\boldsymbol{G}\boldsymbol{\delta}_j\right)}\left(F\left(\boldsymbol{G}'\boldsymbol{G}\boldsymbol{\delta}_{\text{ref}}\right)+R_\omega\right) \\ -\bar{\lambda}_j^2 \sqrt{F\left(\boldsymbol{G}'\boldsymbol{G}\boldsymbol{\delta}_{\text{ref}}\right)} F\left(\boldsymbol{G}'\boldsymbol{G}\boldsymbol{\delta}_j\right). \quad (4.41)$$

By defining $B_\omega^{\text{ref}} \triangleq F\left(\boldsymbol{G}'\boldsymbol{G}\boldsymbol{\delta}_{\text{ref}}\right)$ and $B_\omega^j \triangleq F\left(\boldsymbol{G}'\boldsymbol{G}\boldsymbol{\delta}_j\right)$, (4.41) can be rewritten as follows:

$$\kappa_j^2 \sqrt{B_\omega^{\text{ref}}} R_\omega \approx \left(\frac{\hat{\lambda}_j \bar{\lambda}_{\text{ref}}^2}{\hat{\lambda}_{\text{ref}}}\right)\sqrt{B_\omega^j}\left(B_\omega^{\text{ref}}+R_\omega\right)-\bar{\lambda}_j^2\sqrt{B_\omega^{\text{ref}}}B_\omega^j. \quad (4.42)$$

We design $\{\kappa_j\}$ by solving the following least squares problem

$$\hat{\kappa}_j = \arg\min_{\kappa_j}\left\|\kappa_j^2\sqrt{B_\omega^{\text{ref}}}R_\omega-\left[\left(\frac{\hat{\lambda}_j\bar{\lambda}_{\text{ref}}^2}{\hat{\lambda}_{\text{ref}}}\right)\sqrt{B_\omega^j}\left(B_\omega^{\text{ref}}+R_\omega\right)-\bar{\lambda}_j^2\sqrt{B_\omega^{\text{ref}}}B_\omega^j\right]\right\|^2. \quad (4.43)$$

The solution to (4.43) is given by

$$\hat{\kappa}_j = \frac{1}{\left\|\sqrt{B_\omega^{\text{ref}}}R_\omega\right\|}\left[\left(\frac{\hat{\lambda}_j\bar{\lambda}_{\text{ref}}^2}{\hat{\lambda}_{\text{ref}}}\right)\langle\sqrt{B_\omega^{\text{ref}}}R_\omega,\sqrt{B_\omega^j}R_\omega\rangle \right. \\ \left. +\left(\frac{\hat{\lambda}_j\bar{\lambda}_{\text{ref}}^2}{\hat{\lambda}_{\text{ref}}}\right)\langle\sqrt{B_\omega^{\text{ref}}}R_\omega,\sqrt{B_\omega^j}B_\omega^{\text{ref}}\rangle-\bar{\lambda}_j^2\langle\sqrt{B_\omega^{\text{ref}}}R_\omega,\sqrt{B_\omega^{\text{ref}}}B_\omega^j\rangle\right]^{\frac{1}{2}}. \quad (4.44)$$

Using (4.26), we simplify (4.44) as follows

$$\hat{\kappa}_j = \sqrt{\left(\frac{\bar{\lambda}_{\text{ref}}^2}{\hat{\lambda}_{\text{ref}}}\right)\hat{\lambda}_j + E_\omega\left[\left(\frac{\bar{\lambda}_{\text{ref}}^2}{\hat{\lambda}_{\text{ref}}}\right)\hat{\lambda}_j-\bar{\lambda}_j^2\right]}, \quad (4.45)$$

where $E_\omega = \langle\sqrt{B_\omega^{\text{ref}}}R_\omega,\sqrt{B_\omega^{\text{ref}}}B_\omega^{\text{ref}}\rangle/\left\|\sqrt{B_\omega^{\text{ref}}}R_\omega\right\|^2$.

Regularizer (2.17) with (4.45) (hereafter N-REG) provides approximately uniform noise properties in the reconstructed image. The new $\{\kappa_j\}$ factors (4.45) consist of two terms within a square root. If we ignore the second term, then (4.45) is approximately the square root of the pre-tuned

Table 4.1: Acronyms for regularizers

Acronym	Description
A-REG	aggregated certainty (AC) regularizer (4.1)
R-REG	regularizer for spatial resolution uniformity (4.27)
N-REG	regularizer for noise uniformity (4.45)

spatial strength (4.27) of R-REG. This suggests that N-REG has less spatial variation of the regularization strength compared to R-REG. The denominator in (4.36) will have the same effect as the denominator in (4.27). Second term of (4.45) is kind of an "adjustment" that is usually smaller than the first term.

4.3 Phantom simulation and clinical experiment

This section investigates the effect of the proposed regularizers (see Table 4.1 for acronyms) for PWLS image reconstruction of 3D short-scan axial CT using both phantom and clinical data. For the hypothetical geometry $\boldsymbol{G}$, we assumed $c_i = 1, \forall i$, and $s_j = 1, \forall j$, in (4.9) and used a full 360° scan with increased number of detector rows, $N_{\mathrm{t}}^{\mathrm{ext}}$ (see Appendix B for a different choice of $\boldsymbol{G}$).

4.3.1 Resolution uniformity

4.3.1.1 Phantom simulation

An anthropomorphic phantom simulation was used to demonstrate the improved spatial resolution uniformity induced by the regularizer with the pre-tuned spatial strength (4.27). We used $512 \times 512 \times 122$ XCAT phantom [112] with voxel size $\Delta_{\mathrm{x}} = \Delta_{\mathrm{y}} = 0.9766$ mm and $\Delta_{\mathrm{z}} = 0.625$ mm as our true image, $\boldsymbol{x}^{\mathrm{true}}$ (Fig. 4.2).

We used the separable footprint projector [79] to simulate a monoenergetic, noiseless sinogram for a 3rd-generation axial cone-beam CT system having $N_{\mathrm{s}} = 888$ channels and $N_{\mathrm{t}} = 64$ detector rows with spacings $\Delta_{\mathrm{s}} = 1.0239$ mm and $\Delta_{\mathrm{t}} = 1.09878$ mm. We assumed a short-scan protocol that covers an angular range of 227.6° with $N_{\mathrm{a}} = 622$ evenly spaced views. We selected the hypothetical geometry $\boldsymbol{G}$ to have both extended views, $N_{\mathrm{a}}^{\mathrm{ext}} = 984$ over 360°, and detector rows, $N_{\mathrm{t}}^{\mathrm{ext}} = 168$. The statistical weights were $w_i = b_0 \exp(-[\boldsymbol{A}\boldsymbol{x}]_i)$ where the X-ray intensity was $b_0 = 10^6$.

To obtain local impulse responses at various locations, we added impulses with amplitude $\epsilon = 2.5 \times 10^{-4}$ mm^{-1}, corresponding to approximately 14 HU, to 6 different locations in each of the selected 9 slices (see Fig. 4.2 for impulse locations in xy plane). Selected slices were evenly spaced through z-dimension including isoplane, end slices of region-of-interest (ROI), and slices

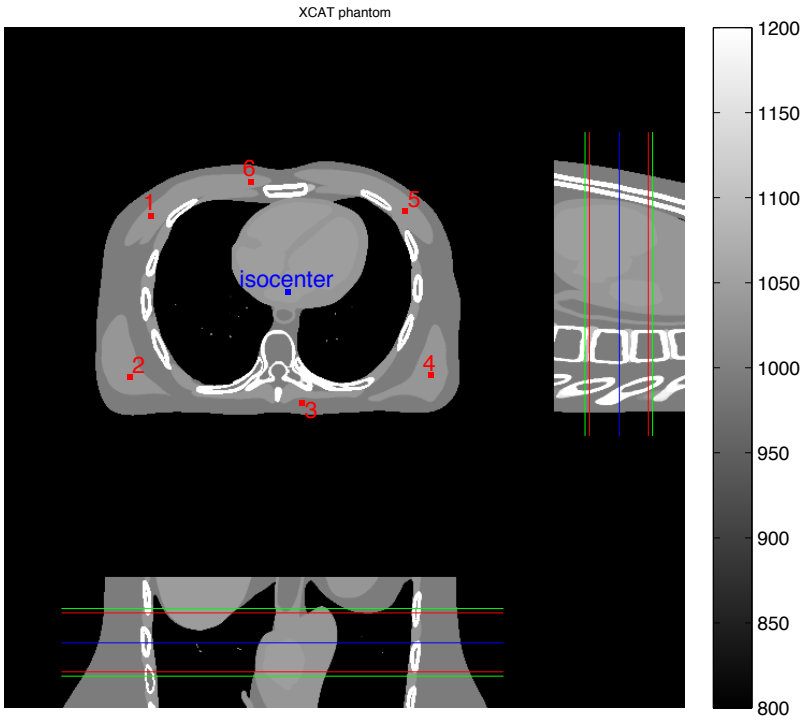
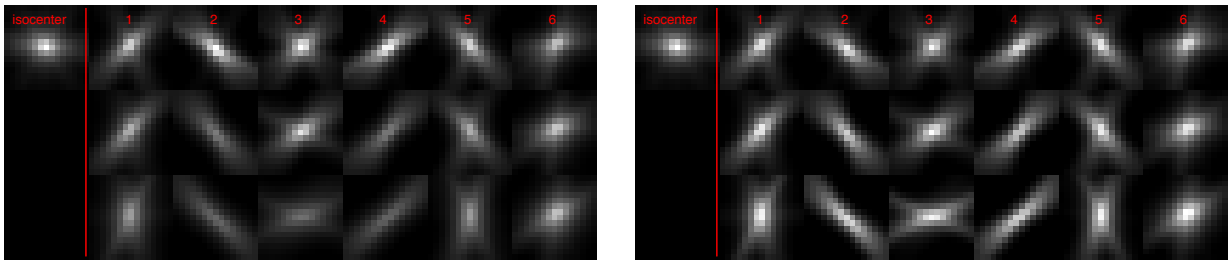


Figure 4.2: XCAT phantom used in the simulation. Middle 3 planes (xy, xz, and yz planes through the isocenter) are shown. Red and blue dots indicate locations of the added impulses and the isocenter, respectively. Red lines indicate the 5th and 56th slices (out of 64 slices), which is our ROI. Blue line displays the location of the center slice. Finally, green lines show the 1st and 64th slices, which is the axial coverage of the simulated scanner.

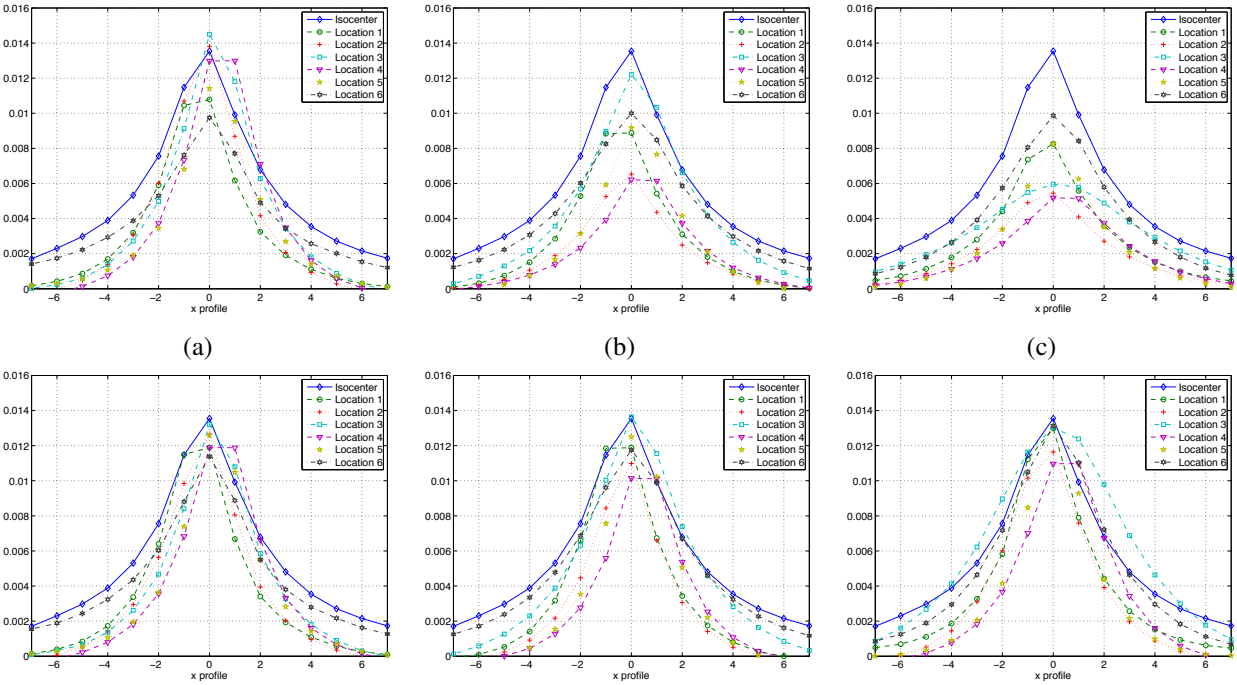
outside ROI. Axial ROI was selected as 5th to 56th slices (out of 64) to avoid displaying slices with possible short-scan artifacts due to insufficient sampling. (Ideally, we want to display the entire coverage, which requires efficient short-scan artifact removal. See Chapter V.) We used (3.6) to evaluate the local impulse response at each location for regularizer designs with both the original aggregated certainty (4.1) (hereafter A-REG) and the pre-tuned spatial strength function (4.27) (R-REG). Both quadratic and edge-preserving regularizations were investigated to show that the proposed regularizer design R-REG is applicable to both cases. Image reconstruction was done on the same grid as the true image. The regularization parameter β was selected based on the full-width at half-maximum (FWHM) of the local impulse response at the isocenter. To visualize the shape of the local impulse response more clearly, we selected a somewhat large β value for which the FWHM was approximately 3 times the voxel size. For this experiment, we set $r^l = 1, \forall l$. Image reconstruction used the ordered-subsets with double surrogates (OSDS) method [17]. The number of iterations was 20 with 41 subsets.

First, we present the results for regularization with a quadratic potential function. Fig. 4.3



(a) A-REG (b) R-REG

Figure 4.3: Comparison of xy plane through the center of each local impulse response at selected location (see Fig. 4.2 for the index of locations). Quadratic potential function was used. Top row is from a center slice (blue line in Fig. 4.2), middle row is from 1st slice of ROI (red line in Fig. 4.2), and bottom row is from outside ROI (green line in Fig. 4.2). (a) Regularization with original aggregated certainty (4.1) (A-REG) (b) Regularization with proposed pre-tuned spatial strength (4.27) (R-REG).



(a) (b) (c)

(d) (e) (f)

Figure 4.4: Comparison of x profiles through the center of each impulse response in Fig. 4.3. Left column is from a center slice, middle column is from 1st slice of ROI, and right column is from outside of ROI. Top and bottom rows represent the regularizers A-REG (4.1) and the proposed R-REG (4.27), respectively. (a) A-REG, center slice (b) A-REG, 1st slice of ROI (c) A-REG, outside ROI (d) R-REG, center slice (e) R-REG, 1st slice of ROI (f) R-REG, outside ROI.

illustrates that the proposed regularizer (4.27) leads to local impulse response functions having more uniform CRC values than the "conventional" aggregated certainty design (4.1), particularly for off-center slices. CRC values of local impulse responses at different locations become nonuniform when using A-REG (4.1). This nonuniformity becomes severe as we move away from center slices. Using the proposed regularizer R-REG (4.27), the CRC values become much more uniform regardless of the location or the amount of sampling. R-REG (4.27) corrects only the nonuniformity of peak values of local impulse responses. Anisotropy in the shape of the impulse response could be improved by designing directional coefficients [19].

Fig. 4.4 shows x profiles through the center of all local impulse responses to compare CRC values more closely. Using A-REG (4.1) leads to resolution nonuniformity even in the center slice, primarily due to short-scan geometry. Nonuniformity in resolution becomes most severe for locations 2 to 4 that have much less sampling compared to the isocenter due to both axial cone-beam geometry and short-scan orbit.

Table 4.2 and Table 4.3 compare the average "mismatch" of the CRC values for given 6 locations across slices and within each slice. We used the following definition of CRC mismatch

$$\rho_j \triangleq \left|\mathrm{CRC}(\boldsymbol{l}_j) - \mathrm{CRC}(\boldsymbol{l}_{\mathrm{ref}})\right| / \left|\mathrm{CRC}(\boldsymbol{l}_{\mathrm{ref}})\right| . \tag{4.46}$$

The proposed regularizer R-REG improved the uniformity of CRC values throughout the reconstruction volume. The average CRC mismatch was significantly improved for all locations and slices, and undersampled voxels were most improved by the proposed regularizer, as designed. The overall improvement of CRC mismatch was from 34.5% to 9.9%.

We obtained similar results for edge-preserving regularization with a hyperbola potential function (2.18). Shape of the local impulse responses does not change much compared to the quadratic regularization, but CRC values become slightly higher. As in the quadratic case, the original certainty function (4.1) yields non-uniform CRC values across multiple voxel locations, and proposed regularization (4.27) leads to more uniform CRC values. Due to their similarity to the quadratic case, results for edge-preserving regularization are presented in Appendix B.

Even though proposed designs were based on approximations, such as (4.15), the local impulse response calculated by (3.6) yields CRC values that closely match the target CRC. For both quadratic and edge-preserving regularizers, the proposed designs provide improved uniform CRC values (see also Table 4.2 and Table 4.3).

4.3.1.2 Real clinical data

We reconstructed a clinical cardiac CT scan as a $1024 \times 1024 \times 122$ image with 70 cm field-of-view (FOV). Measurements were obtained from a 64 row axial CT scanner with a short-scan

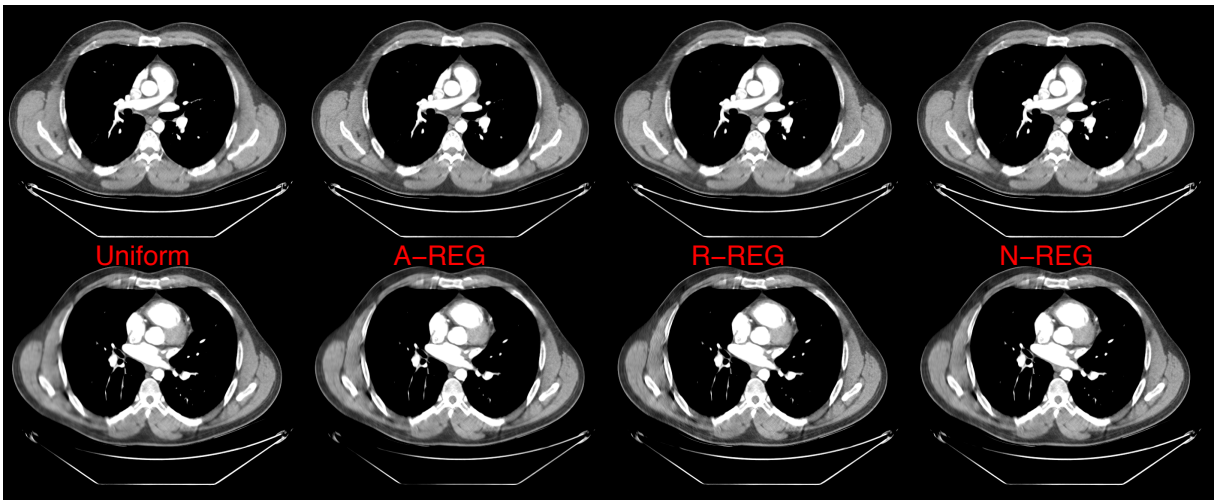


Figure 4.5: Reconstructed images using uniform regularizer (1st column), A-REG (2nd column), R-REG (3rd column), and N-REG (last column). Quadratic potential function was used. Top row: center slice; bottom row: the last slice of ROI. Display range is [800 1200] (HU).

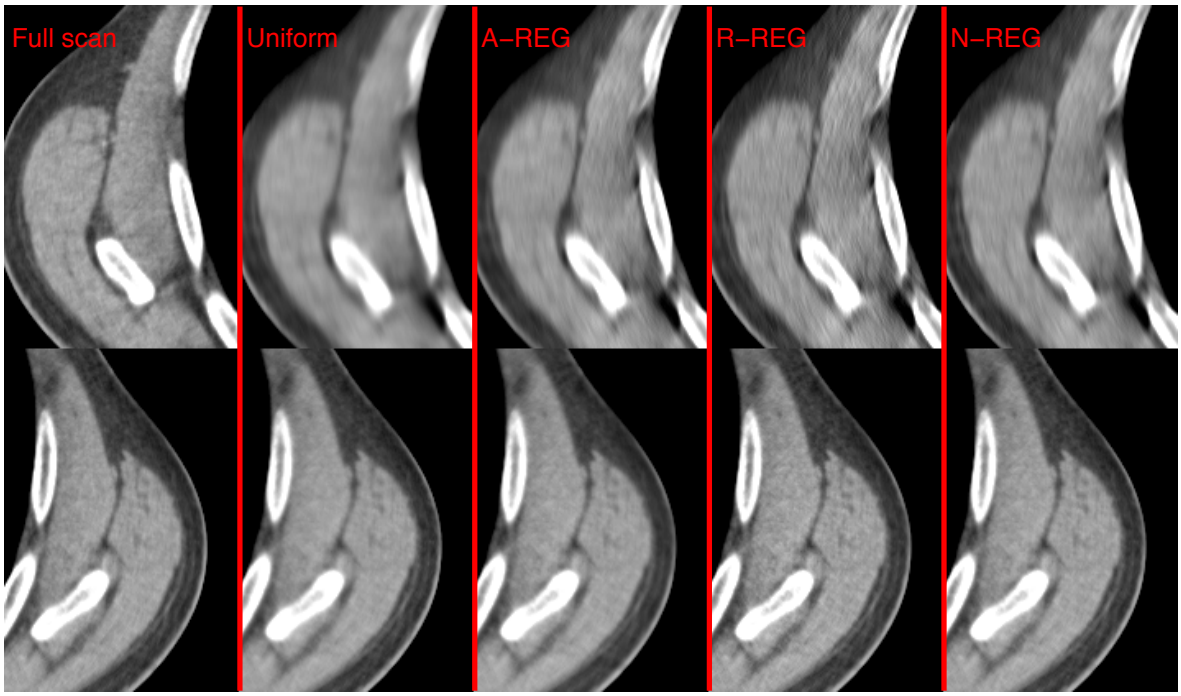


Figure 4.6: Comparison of reconstructed images in Fig. 4.5 at the last slice of ROI. From left to right, the images are from full scan measurements with A-REG, uniform regularizer, A-REG, R-REG, and N-REG, respectively. Top row is from a region on the left side where sampling is lower than a region on the right side (bottom row). Display range is [800 1200] (HU).

protocol and 480 mAs tube current. The sinogram dimension was $[N_s, N_t, N_a] = [888, 64, 642]$. We selected the hypothetical geometry $\boldsymbol{G}$ to have both extended views, $N_{\mathrm{a}}^{\mathrm{ext}} = 984$, and detector rows, $N_{\mathrm{t}}^{\mathrm{ext}} = 148$. For this experiment, we set r^l using [138]. We used ICD with spatially non-homogeneous updates [140] for reconstruction. We show results from both quadratic and edge-preserving regularization, for which we used the q-generalized Gaussian potential function with $p = 2$, $q = 1.2$, and $c = 10$, [124]

$$\phi(t) = \frac{|t|^p}{1 + |t/c|^{p-q}}. \tag{4.47}$$

Fig. 4.5 compares the reconstructed images with a quadratic potential function and the following different regularizers: Uniform, A-REG, and R-REG. When uniform regularization is used, *i.e.*, $\kappa_j = 1, \ \forall j$, the reconstructed image becomes over-smoothed even for some locations in the center slice, illustrating the importance of the pre-tuned spatial strength in the regularization. Reconstructed image using A-REG (4.1) shows less blurring and sharper spatial resolution compared to that of the uniform regularizer. However, even in the center slice, both left and right sides of the reconstructed image have different resolution. This is consistent with the result in Fig. 4.4 (a) where the CRC values were nonuniform even in the center slice due to short-scan orbit. The proposed regularizer R-REG (4.27) improved resolution uniformity in the center slice. In the end slices of the ROI, the resolution nonuniformity becomes more apparent. A-REG (4.1) fails to provide resolution uniformity at under-sampled locations, leading to visible differences in smoothness between left and right side of the reconstructed images. On the other hand, the reconstructed image using R-REG (4.27) has more uniform resolution properties even in these undersampled region, leading the structures in the region to have more sharp and clear boundaries.

Fig. 4.7 and Fig. 4.8 show reconstructed images for edge-preserving regularization. Clearly, the edge-preserving regularization preserves fine structures, leading to better image quality in terms of spatial resolution compared to the quadratic regularization. However, the choice of the regularizer still affects the resolution uniformity in the reconstructed image. The results show similar tendency as in quadratic case: non-uniform resolution and over-regularization in undersampled region for uniform regularizer and A-REG. On the other hand, the proposed R-REG achieves sharper and more uniform spatial resolution. This suggests that even though the proposed regularizer was derived for a quadratic regularization, it can also be used well for non-quadratic regularization.

4.3.2 Noise uniformity

Reconstructed images using R-REG (4.27) shows better resolution uniformity throughout the entire volume, however, this improvement in the spatial resolution was achieved at the expense of the noise properties in the image. As shown in Fig. 4.8, the proposed R-REG slightly increases the noise level in the reconstructed image when edge-preserving regularization is used. This trade-

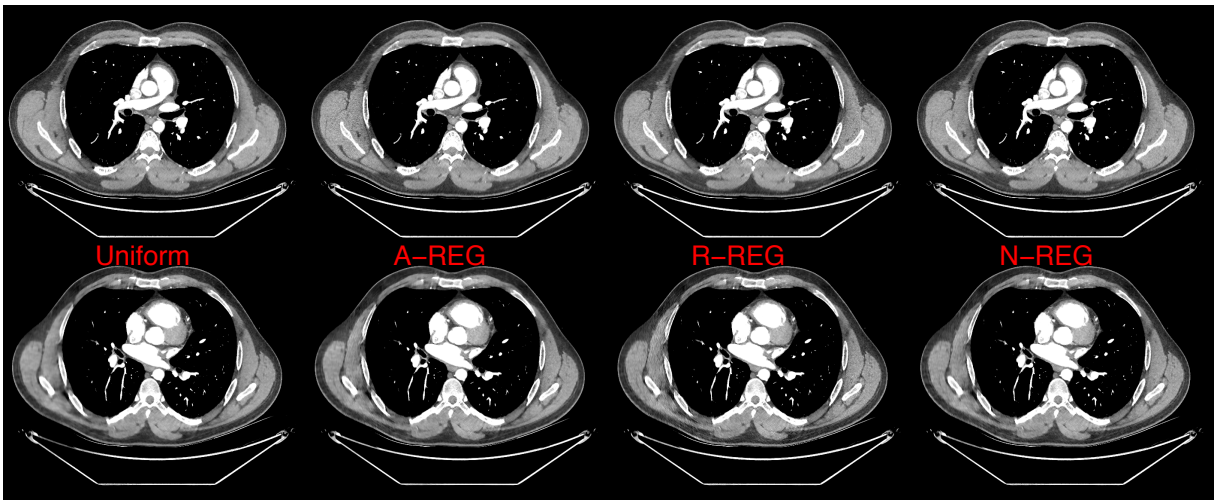


Figure 4.7: Reconstructed images using uniform regularizer (1st column), A-REG (2nd column), R-REG (3rd column), and N-REG (last column). Edge-preserving potential function (4.47) was used. Top row: center slice; bottom row: the last slice of ROI. Display range is [800 1200] (HU).

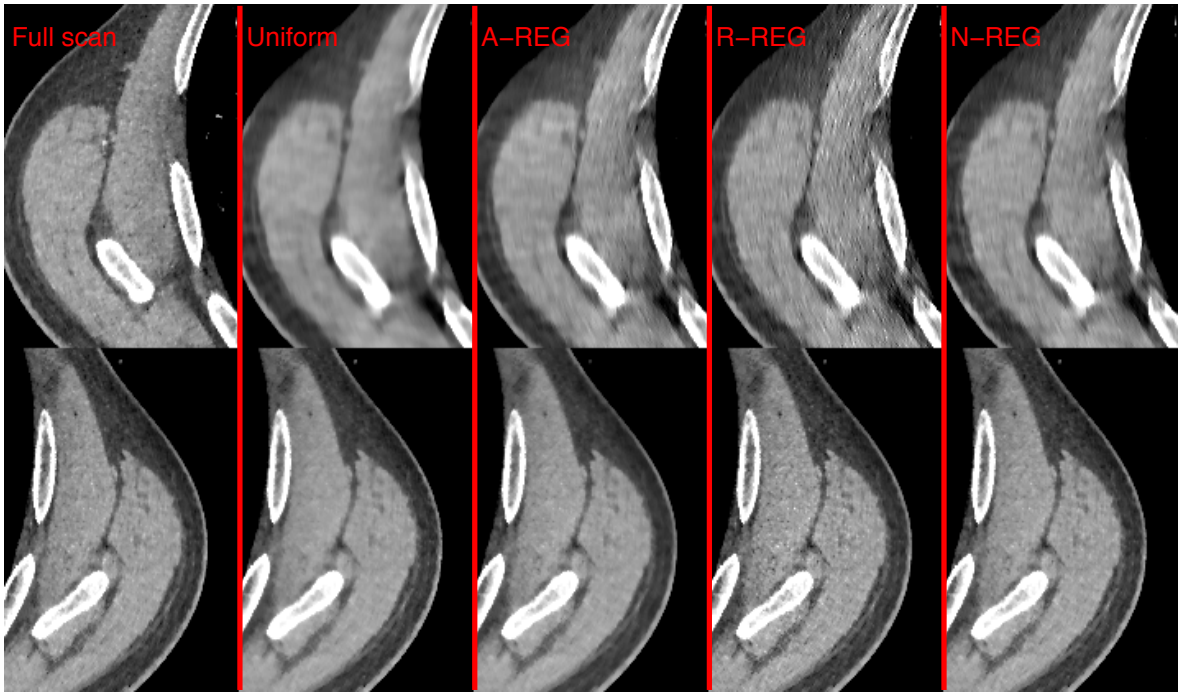


Figure 4.8: Comparison of reconstructed images in Fig. 4.7 at the last slice of ROI. From left to right, the images are from full scan measurements with A-REG, uniform regularizer, A-REG, R-REG, and N-REG, respectively. Top row is from a region on the left side where sampling is lower than a region on the right side (bottom row). Display range is [800 1200] (HU).

off is inevitable, thus we also investigated a regularizer that focuses on noise uniformity (N-REG, (4.45)).

4.3.2.1 Phantom experiment

To compare the regularizers quantitativly, we use the GE performance phantom (GEPP) [56]. The phantom consists of a Plexiglas$^{\text{TM}}$ insert with resolution bars, and tungsten wires in water. The phantom was scanned with a 64 row axial CT scanner in short-scan mode and 70 mAs tube current, which corresponds to very low dose scan, and reconstructed to a grid of $1024 \times 1024 \times 90$ with the following voxel size: $\Delta_x = \Delta_y = 0.2246$ mm and $\Delta_z = 0.625$. We selected the scanning geometry with both extended views, $N_{\text{a}}^{\text{ext}} = 984$, and detector rows, $N_{\text{t}}^{\text{ext}} = 200$, as the hypothetical geometry $\boldsymbol{G}$. Edge-preserving regularization with q-generalized Gaussian potential function (4.47) was used. We set r^l using [138].

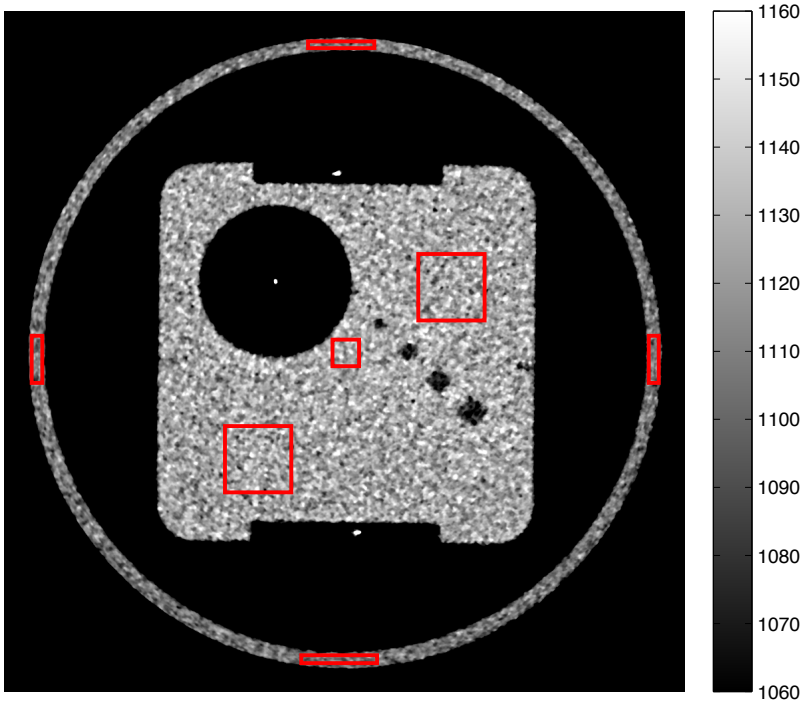


Figure 4.9: The GEPP used for quantitative comparison of regularizations. Red boxes indicate the regions selected for noise variance comparison.

Fig. 4.9 shows the reconstructed image of the GEPP with previous regularization. Due to small FOV (= 23 cm), the end slices of ROI did not suffer much from under-sampling. However, the choice of regularization still leads to different image qualities in reconstructed images even in the center slice. To compare different regularizations fairly, we chose the regularization parameter β such that the noise standard deviation near the isocenter is similar for all reconstructed images

Table 4.2: Average CRC mismatch (4.46) for selected 6 locations across slices (Units: %). See Fig. 4.2 for the index of impulse locations.

	Locations					
	1	2	3	4	5	6
A-REG (Fig. 4.4)	32.6	45.3	24.8	47.2	31.9	25.0
R-REG (Fig. 4.4)	7.5	14.9	3.9	20.0	5.9	7.2

Table 4.3: Average CRC mismatch (4.46) within each slice, respectively (Units: %).

	Averages			
	Overall	Center Slice	1st slice of ROI	Outside ROI
A-REG (Fig. 4.4)	34.5	(a) 12.3	(b) 33.7	(c) 46.8
R-REG (Fig. 4.4)	9.9	(d) 8.7	(e) 11.6	(f) 7.4

(≈ 13.7 HU). We selected 7 different homogeneous regions in the center slice to compare the noise standard deviation (see Fig. 4.9 for their locations). Table 4.4 illustrates that the proposed N-REG shows the best noise uniformity out of given regularizers, *i.e.*, the average standard deviation of the noise is reasonably close to that of the region near isocenter. Since FOV is small, the noise standard deviation does not vary much within Plexiglas™ insert. However, the standard deviation in the wall shows noticeable variability depending on the regularization method. Due to the symmetrical shape of the GEPP and thickness of its wall, the statistical weighting varies mostly only in channel dimension (except for the views affected by an additional weighting such as Parker weighting) and has an "U" shape. As a result, uniform regularizer generally increased noise in the reconstructed image, and both A-REG and R-REG over-regularized the region far away from isocenter, *i.e.*, walls in this case. Both A-REG and R-REG showed similar performance as expected, and N-REG improved noise uniformity in the reconstructed image.

4.3.2.2 Real clinical data

Fig. 4.5 to Fig. 4.8 compare the reconstructed images obtained using the proposed N-REG with other regularizers. For both quadratic and edge-preserving regularizations, the proposed N-REG provides improved noise uniformity in the reconstructed image compared to other regularizers. Uniform regularizer tends to over-smooth the reconstructed image, and A-REG shows nonuniform noise properties even in the center slice. Both proposed regularizers R-REG and N-REG show improved improved image qualities in terms of both resolution and noise. R-REG shows somewhat shaper reconstructed image compared to N-REG on end slices for both regularizations, but has slightly higher and nonuniform noise variance.

Fig. 4.6 and Fig. 4.8 zoom into the reconstructed images at the last slice of ROI, which has nonuniform sampling within the slice. A reconstructed image from full scan measurements with A-

REG was used as a reference with desirable image quality. Since the image quality of the reference image is also affected by the choice of regularization, it may not be the optimal image for clinical diagnosis. However, the under-sampling from short-scan measurements is a more dominant factor for the image quality in the displayed region, so the chosen reference image shows better image characteristics compared to other reconstructed images from short-scan measurements. For each case, the region on the right side of the reconstruct image (bottom row) was compared to the left side of the image that has less sampling (top row). Uniform regularization clearly leads to over-regularization in the undersampled region, yielding severe noise nonuniformity within the slice. Even though less severe, A-REG also suffers from the same issue. Proposed regularization N-REG generates reconstructed image with better noise uniformity, leading the structures in the undersampled region to have more clear visibility.

For both examples, the proposed N-REG (4.45) provides more uniform noise characteristics in the reconstructed image compared to other regularization methods.

Table 4.4: Comparison of the noise standard deviation for different regularizers method. All values are in Hounsfield units (HU).

	Uniform	A-REG	R-REG (4.27)	N-REG (4.45)
Plexiglas^{TM} Insert (center)	14.0	13.7	13.7	13.6
Plexiglas^{TM} Insert (right)	18.8	16.0	16.2	16.4
Plexiglas^{TM} Insert (left)	15.1	14.4	14.5	14.9
Phantom Wall (left)	15.4	9.3	9.5	11.8
Phantom Wall (right)	16.9	8.8	6.8	10.1
Phantom Wall (top)	17.4	8.7	8.8	10.7
Phantom Wall (bottom)	16.1	7.7	8.8	10.9
Average	16.2 ± 1.6	11.2 ± 3.4	11.2 ± 3.6	12.6 ± 2.4

4.4 Discussion

We have proposed new regularization methods by modifying the aggregated certainty presented in [42] using the hypothetical geometry concept. Proposed regularization (4.27) improved the spatial resolution uniformity in the reconstructed image, and (4.45) provided better uniform noise characteristics compared to uniform and aggregated certainty regularization methods.

Proposed methods, R-REG and N-REG, showed improved spatial resolution or noise uniformity compared to uniform regularizer and A-REG in both quadratic and edge-preserving regularizations and for both simulated and clinical scans. Even though the proposed regularizers were targeted to improve the uniformity of either the spatial resolution or noise, they yield reconstructed images with qualitatively improved image quality in terms of both resolution and noise compared

to that from the uniform and the aggregated certainty regularizers. For quadratic regularization, the noise characteristics have less effect on the visual image quality than the spatial resolution, which suggests the use of R-REG to improve spatial resolution uniformity in the reconstructed image. On the other hand, since edge-preserving regularization provides improved resolution by enhancing sharply changing structures, the noise uniformity primarily affects the readability of the reconstructed image. Thus, N-REG may be preferable for edge-preserving regularization. However, there exists trade-off between spatial resolution and noise characteristics. Using either regularizer may not provide an optimal reconstructed image in terms of both resolution and noise. Diagnostic readability for the reconstructed images obtained from both methods needs to be carefully investigated to determine the best regularization, and possibly some combination of methods may be desirable. A compromise regularization that balances spatial resolution and noise characteristics is explored in the supplement as a starting point for further research.

For experiments in the chapter, we used the hypothetical scanning geometry obtained intuitively by extending both rows and views from given axial cone-beam CT geometry. We also suggested other possibilities such as step-and-shoot set of axial scans. However, for other geometry, determining the appropriate hypothetical geometry may be harder. For example, in helical CT, simply extending views would not suffice, and since the actual scanning geometry must be a subset of the hypothetical geometry, we cannot use a very small pitch. Intertwined multiple helical geometries may be a possible choice. As shown in this example, a careful consideration is required to properly extend the proposed regularizations to other scanning geometries.

One minor drawback of using the generalized geometry is the increased computation for (4.12). For the geometry with extended views and rows, since (4.12) is calculated only once prior to the iterating process, the increase in the computation expense is insignificant compared to the computation required for the actual reconstruction. However, the use of step-and-shoot set of axial scans or intertwined multiple helical scans may require considerable computations. Fortunately, since the calculation of (4.12) only requires the information of the hypothetical geometry and the same hypothetical geometry can be used for different geometries as long as they are subsets, one could tabulate the denominator part of (4.12).

Proposed regularizers improve the uniformity of spatial resolution or noise by controlling a scaling factor at each voxel. Even though the design process involves matching the entire local impulse responses and noise correlation functions, they are primarily matching CRC values and variances at each location due to the approximation (4.15) and (4.37). Thus, proposed regularizers do not correct anisotropy of these characteristics. Designing directional coefficients in the regularization may have potential to correct for such anisotropy, and have shown promising results for the well-sampled regions [20, 116]. However, the anisotropy of the image characteristics in the under-sampled region is hard to correct, especially since these methods rely on Fourier transform

with locally circulant approximation for Fisher information matrix. One possible future work is to extend the methods in this chapter to directional coefficient design.

CHAPTER V

Short-scan artifact removal 3D axial cone-beam CT using all available information

As described in Chapter II, the short-scan approach has several advantages compared to the full-scan approach due to its reduced scan time, but suffers from insufficient sampling at certain voxel locations. Using more measurements for reconstruction reduces such artifacts, but at the expense of decreased temporal resolution. This chapter investigates methods that reduce short-scan artifacts while preserving the temporal resolution in the well-sampled slices.

We fist propose a statistical weighting modification approach that controls the contribution of the weights associated with the extra views available. We then introduce another regularization term to the conventional statistical image reconstruction cost function to obtain structural information from an initial reconstruction.

5.1 Short-scan artifact removal using all acquired views

Clinical cardiac scans typically include extra views for the purpose of viewing different phases, motion-compensation algorithms, etc.. Here we use such extra acquisition data to reduce short-scan artifacts in the undersampled region. Using XCAT phantom simulations, we investigated an approach to modifying the statistical weights associated with the extra views. This approach reduced short-scan artifacts, particularly in slices near the ends of the axial field-of view (FOV).

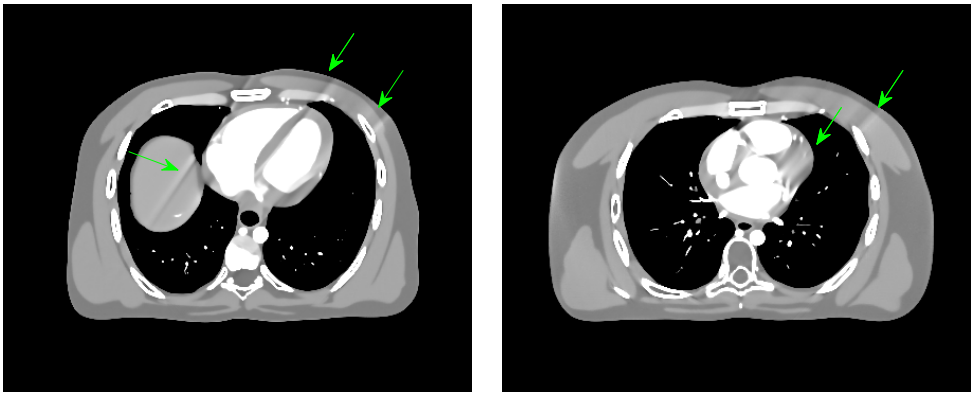
5.1.1 Introduction

5.1.1.1 Motivation

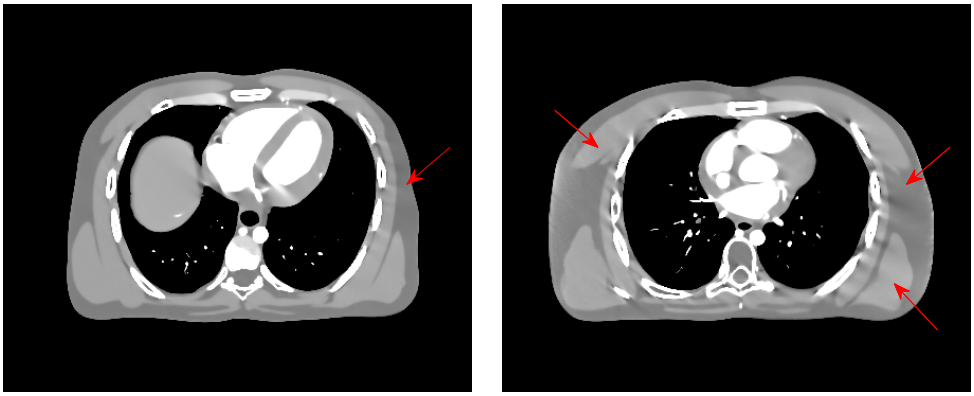
Short-scan approach leads to insufficient sampling at certain voxels, and more voxels in the slice become insufficiently sampled as one moves away from the iso-plane. This situation becomes

This chapter is based on [23]

severe as the cone angle becomes larger. Using more projection views for reconstruction reduces such artifacts, but at the expense of decreased temporal resolution. Reconstructed images of the XCAT phantom using both full-scan and short-scan measurements are shown in Fig. 5.1. A 64-slice CT scanner was simulated, and the images at 1st and 64th slices were displayed where regular cardiac images use slices between 5th to 56th slices for diagnosis. Fig. 5.1 (b) shows that, for short-scan, as we move toward end slices, we see noticeable artifacts (indicated by red arrows) in the reconstructed images resulting from insufficient sampling. On the other hand, reconstructed images from full-scan measurements (Fig. 5.1 (a)) do not show such artifacts, but suffer from severe motion artifacts (indicated by green arrows).



(a) Full-scan, end slices



(b) Short-scan, end slices

Figure 5.1: Reconstructed images of the XCAT phantom using full-scan measurements and short-scan measurements at end slices. Measurements were obtained from a 64-slice CT scanner with 40 mm collimation. Green and red arrows indicate motion artifacts and short-scan artifacts, respectively. Display window is [850 1150] (HU).

Current image reconstruction methods do not provide satisfactory images for short-scans at those slices far away from the iso-plane. We want to develop a method that reduces short-scan artifacts on those slices. In clinical settings, it is common to acquire extra measurements in view dimension for purposes such as viewing different phases, use of motion-compensation algorithms,

etc.. In this study, we investigated a method for statistical image reconstruction that uses such extra acquisition data to reduce short-scan artifacts in the undersampled region. The goal is to use the part of extra measurements to remove short-scan artifacts without compromising the temporal resolution in the sufficiently sampled region, or so called "football" region. We propose a statistical weighting modification approach that was evaluated with the XCAT phantom simulation.

5.1.1.2 Sinogram Extrapolation

Before introducing the statistical weighting approach, let us elaborate about "extra" measurements. As mentioned earlier, in clinical settings, it is common to acquire extra measurements in view dimension for various purposes. Such measurements, when used in the reconstruction, can help reduce short-scan artifacts and improve SNR at the expense of degraded temporal resolution. What if we use data extrapolation to obtain such extra measurements without increasing the scan time or the number of detector rows? In practice, FBP reconstruction uses data extrapolation for the reconstruction of undersampled region. Zeroth order extrapolation in the detector row dimension is used to extend its axial coverage. Surprisingly, this works well, and generates reasonable images. Motivated by the success in FBP reconstruction, we evaluated sinogram extrapolation in both view and row dimension.

Sinogram extrapolation in row dimension For the measurement $\boldsymbol{y} \triangleq y(s,t,v)$ where s, t, and v denote channel, row, and view dimensions respectively, the row extrapolated measurement $\boldsymbol{y}_{\text{ext}} \triangleq y_{\text{ext}}(\cdot,\cdot,\cdot)$ can be written as

$$y_{\text{ext}}(\cdot,t,\cdot) = \begin{cases} y(\cdot,t,\cdot) & \text{if } t_{min} \le t \le t_{max} \\ y(\cdot,t_{min},\cdot) & \text{if } t \le t_{min} \\ y(\cdot,t_{max},\cdot) & \text{if } t \ge t_{max}, \end{cases} \tag{5.1}$$

where $[t_{min}\ t_{max}]$ is the range of t for real measurements.

Fig. 5.2 shows the results from a preliminary simulation with the XCAT phantom. It added some additional artifacts to the image. Depending on the object, extrapolated measurements may or may not be good approximation.

Furthermore, the statistical weighting was also extrapolated with zeroth order approximation as (5.1), which means that both real and extrapolated data were treated equally. Some down-weighting for the extrapolated rows may lead to better image quality by balancing the effects of real and extrapolated data. Fig. 5.3 shows the results with two down weighting values, 0.25 and 0.01. Both results show increased artifact in the reconstructed image. Some of the structures

may be considered as sharper, but the increased artifacts have more dominant effects on the image quality.

Since there are many parameters involved including number of padded rows, how to weight the padded region, and how to adjust the regularization strength, it is hard to claim that this was a definitive study. Using more sophisticated extrapolation methods may provide better results. However, using a simple method such as zeroth order extrapolation is not desirable for statistical image reconstruction methods, since the extrapolated data are inconsistent with the statistical model that assumes independence.

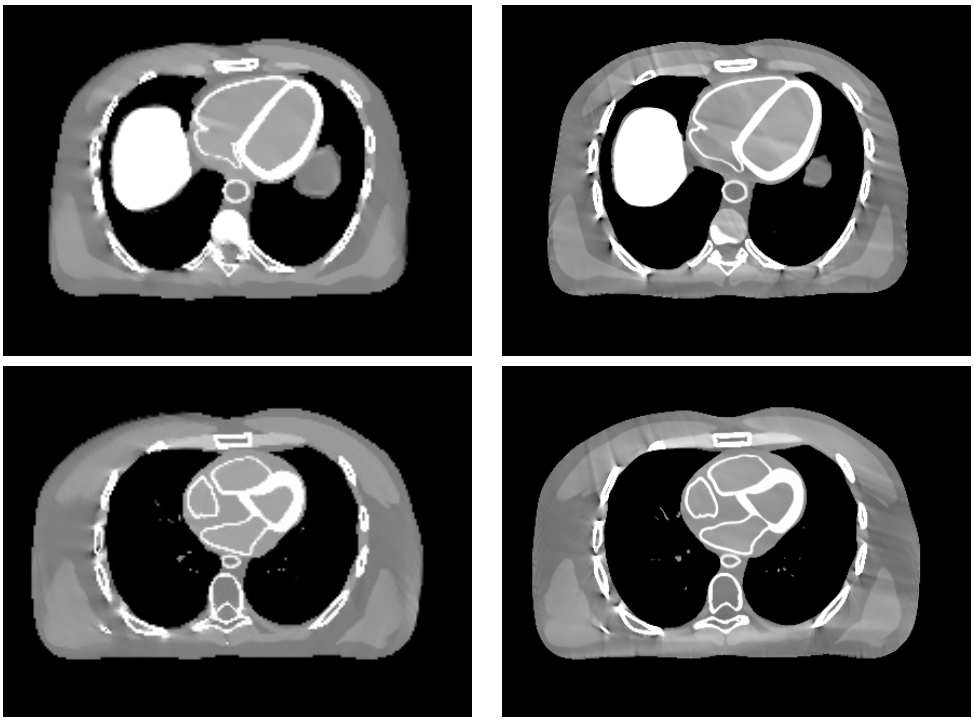


Figure 5.2: Reconstructed images of a XCAT phantom using ordered-subsets method with 82 subsets. A 64-slice CT scanner was simulated. Left column: standard reconstruction. Right column: Zeroth order extrapolation was used to the projection data. Thirty rows were padded to both top and bottom of the rows.

Sinogram extrapolation in view dimension The artifacts come from missing projections due to the geometry of the short-scan. To provide those projections, we re-project an initially reconstructed image using either FBP or a statistical method. Since this initially reconstructed image has artifacts, the sinogram obtained by re-projecting it may not give us a perfect image. But they may reduce the artifacts at the end slice.

If we extrapolate the sinogram in view domain as we did in row dimension (5.1), the voxels with sufficient sampling, such as voxels near the isocenter, will also be affected, and this is not

(a) Down weighting = 0.01

(b) Down weighting = 0.25

Figure 5.3: Short-scan reconstruction with row data extrapolation (a) down weighting = 0.01 (b) down weighting = 0.25. From left to right, each column corresponds to 1st, 17th, center, 239th, last slice of ROI, respectively.

desirable. Therefore, we define the view extrapolated measurement $\boldsymbol{y}_{\mathrm{v,ext}} \triangleq y_{\mathrm{v,ext}}(\cdot,\cdot,\cdot)$ as follows so that only the end slices are affected by the extrapolation:

$$y_{\mathrm{v,ext}}(\cdot,t,v) = \begin{cases} y_{\mathrm{re}}(\cdot,t,v) & \text{if } v \geq v_{\mathrm{short}} \text{ and } t_{min} \leq t \leq t_{max} \\ y(\cdot,t,v) & \text{otherwise,} \end{cases} \tag{5.2}$$

where $\boldsymbol{y}_{\mathrm{re}} \triangleq y_{\mathrm{re}}(\cdot,\cdot,\cdot)$ is the re-projection, v_{short} is the number of views for the short-scan, and $[t_{min}\ t_{max}]$ is the range of t that will not be affected by the extrapolation. The statistical weighting was also extrapolated in the same way.

We obtained the re-projection from the statistical reconstruction of short-scan measurements. Fig. 5.4 shows that the method provides no improvements. The re-projection was obtained from the reconstructed image with artifacts and noises, hence inconsistent with original projection data. This inconsistency corrupts the reconstructed images and induces other artifacts. The results in Fig. 5.4 (d) were obtained by re-projecting the full-scan image, so the artifacts are reduced. These results illustrate the importance of the accuracy of the extrapolated sinogram. The range of t for Fig. 5.4 (d) was selected to show that the range may determine the trade-off between the image quality of the end slices and the inner slices. In this case, the range was not narrow enough to reduce the entire artifacts even with good re-projections.

Similar to previous extrapolation method, this method still has many things to refine such as the range of t and possible down-weighting of the statistical weighting.

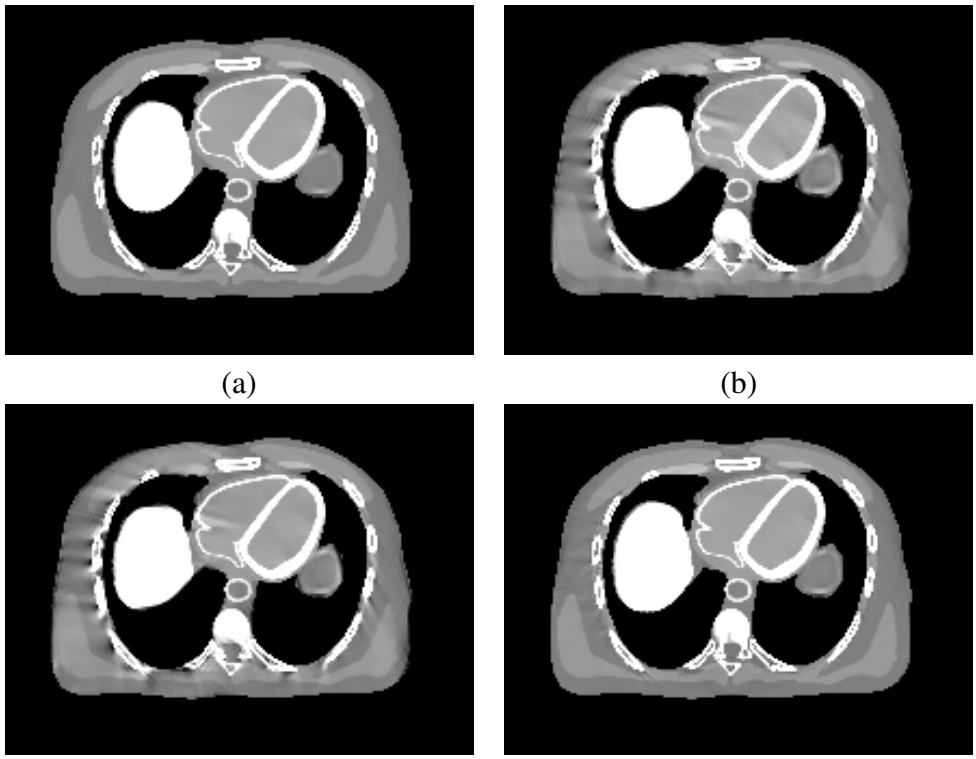


Figure 5.4: Reconstructed images of a XCAT phantom using ordered-subsets method with (a) full-scan measurements, (b) short-scan measurements, (c) view extrapolation where re-projection obtained from (b), and (d) view extrapolation where re-projection obtained from (a). A 64-slice CT scanner was assumed.

5.1.2 Statistical weighting modification approach

Consider a PWLS objective function of the form (3.1). Basic idea behind statistical weighting modification is to selectively use additional measured views for each voxel depending on the sampling properties of the short-scan views. Undersampled voxels can benefit from using the additional measurements. However, some of these additional measurements affect not only the undersampled voxels but also other sufficiently sampled voxels, thus potentially degrading their temporal resolution. At one extreme, using all the additional measured views would cause all voxels to have the deteriorated temporal resolution of the full-scan reconstruction. Therefore, our goal is to optimize the use of additional measurements so that the temporal resolution of the well-sampled reconstructed slices are minimally affected and at the same time the short-scan artifacts are reduced. Since there is a trade-off between the temporal resolution and the short-scan artifact reduction, we need to determine which characteristic is more important for each voxel. For our analysis below, we strongly encourage the voxels inside the football region to have the same temporal resolution as the short-scan reconstruction. For all other voxels, we maximize the use of

additional measurements.

First, we introduce the following new metric to quantify the effect of each detector measurement for a targeted region

$$e_i = \frac{\sum_{j=1}^{n_\mathrm{p}} a_{ij} m_j^{(t)}}{\sum_{j=1}^{n_\mathrm{p}} a_{ij} m_j^{(c)}}, \tag{5.3}$$

where a_{ij} is the elements of the system matrix, n_p is the number of voxels, $m_j^{(t)}$ is the jth element of the mask for a targeted region, and $m_j^{(c)}$ is the jth element of the cylindrical mask. The value of this metric varies from 0 to 1 where 0 means that the ray corresponding to the detector cell does not pass through the targeted region. As mentioned above, we use football region as our targeted region, but one may define the mask differently depending on the purpose.

Second, we modify the statistical weighting based on this metric as follows

$$\hat{\boldsymbol{W}} \triangleq \boldsymbol{W}\boldsymbol{D}[p_i]\boldsymbol{D}[v_i] = \boldsymbol{D}[w_i p_i v_i], \tag{5.4}$$

where $\boldsymbol{W} = \{w_i\}$ is a conventional choice for the statistical weighting, p_i indicates the Parker weighting used to match the temporal resolution of the FBP reconstruction, and

$$v_i = \begin{cases} 1, & \text{if } i \text{ is within short-scan view range} \\ 1 - e_i & \text{otherwise.} \end{cases} \tag{5.5}$$

Fig. 5.5 compares back projections of the following weights: uniform weights with both full and short-scan views, and the designed weight v_i. These back projection images reflect the sampling condition at each voxel. In this figure, Parker weighting was not included for better visualization of the improvements. The designed weight leads to the sampling condition similar to that of the short-scan at the center slice and the full-scan at the end slice, which suggests that our design has potential value.

To add more freedom to our design, we investigate the following design

$$\bar{v}_i = f(v_i) = (v_i)^p, \tag{5.6}$$

where p is a parameter that requires tunning. Fig. 5.6 compares the designed weights along row dimension for different choice of p. As p becomes larger, the use of additional measurements decreases, and the reconstructed image suffers more from the short-scan artifact. On the other hand, small p leads to reduced temporal resolution in the targeted region.

In addition to visual comparison, we need a metric to quantitatively evaluate the reconstructed image obtained by the proposed method. Most intuitive measure would be counting the number of

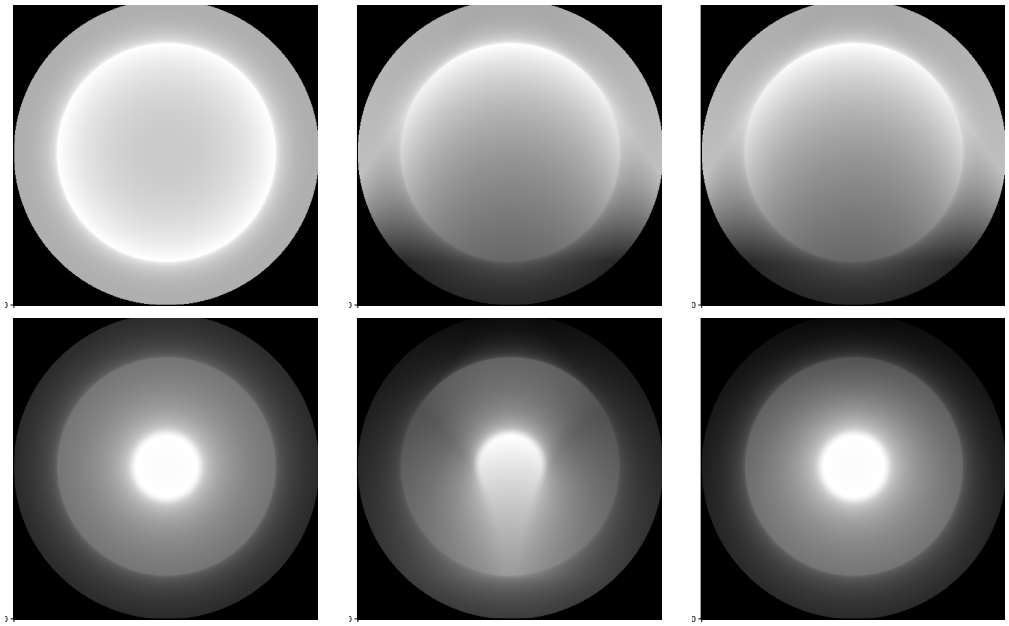
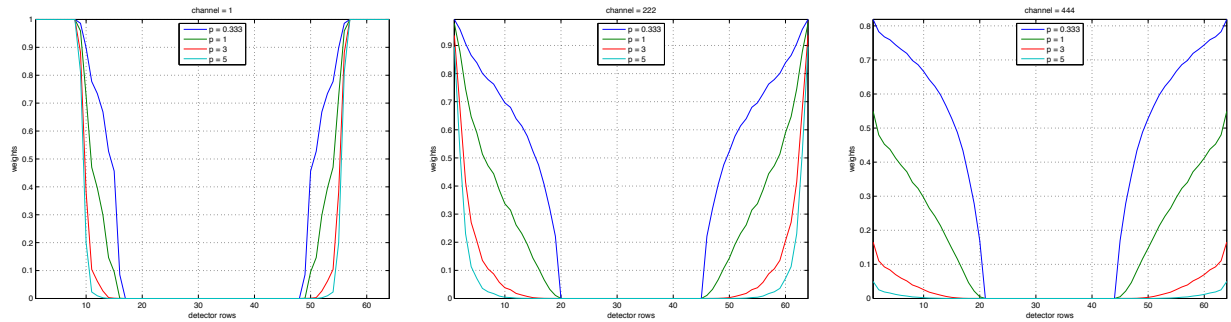


Figure 5.5: Comparison of back projections the following weights: uniform weights with full-scan views (left), uniform weights with short-scan view (middle), and the designed weight (right). Top row is the center slice, and bottom row is the 1st slice of ROI for 64-slice scanner.



(a) (b) (c)

Figure 5.6: Plot of designed weights along row dimension for different p values. (a) 1st channel (b) 222th channel (c) 444th channel.

views or detector elements affecting each voxel when our modified weighting is used. However, calculating such measure may not be trivial or computationally efficient. Instead, we propose the following metric called "view mask" for a given statistical weighting w_i:

$$M \triangleq \sum_n \frac{\sum_{l,m} a_{(l,m,n)j} w_{(l,m,n)}}{\sum_{l,m} a_{(l,m,n)j} o_{(l,m,n)}}, \tag{5.7}$$

where $o_i = o_{(l,m,n)} = 1, \ \forall i$, and l, m, and n indicate channel, row, and view indexes in the

projection domain, respectively. The ratio between the view masks of two different weightings, or "view ratio", can be used to quantify the "closeness" of the proposed method to either full-scan MBIR or short-scan MBIR at certain location. For example, the view ratio between the proposed method and short-scan MBIR is given by

$$R \triangleq \frac{M_{proposed}}{M_{SS}} = \sum_n \frac{\sum_{l,m} a_{(l,m,n)j} \hat{w}_{(l,m,n)}}{\sum_{l,m} a_{(l,m,n)j} \bar{w}_{(l,m,n)}}, \tag{5.8}$$

where $\hat{w}_i = w_i p_i \bar{v}_i$, and $\bar{w}_i = w_i p_i$. For this case, when a voxel has $R = 1$, the temporal resolution of the proposed method is equivalent to that of short-scan MBIR at that location. For our analysis, we want the voxels inside the football region to have the view ratio R close to 1.

5.1.3 XCAT simulations results for statistical weighting modification

The proposed method was investigated on a 3D cone-beam CT imaging problem with 40 mm collimation. The simulated system has 888 channels, 64 detector rows, and 984 evenly spaced view angles over 360°. The XCAT phantom with cardiac motion of 75 bpm was used as an object. To focus on the sampling induced artifacts, we generated noiseless measurements. ICD with spatially non-homogeneous updates [140] was used to reconstruct images. For illustration purposes, we assumed that the acquisition time is long enough so that full-scan measurements is available. However, the proposed method can be applied with any amount of extra acquisition.

Fig. 5.7 shows the view ratio between the proposed method with $p = 5$ and short-scan MBIR within the football mask. The view ratio values at most of the voxels are very close to 1, which indicates that the proposed method provides reconstructed images having the desired temporal resolution in most spatial locations of the targeted region.

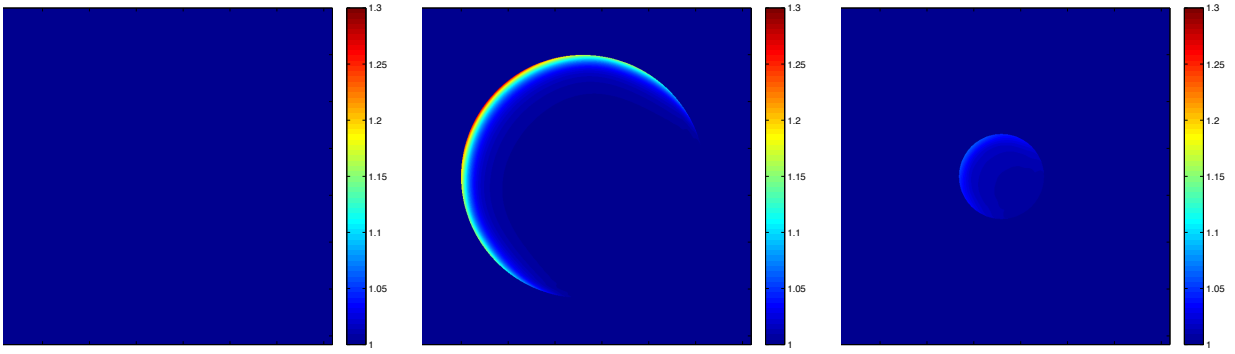
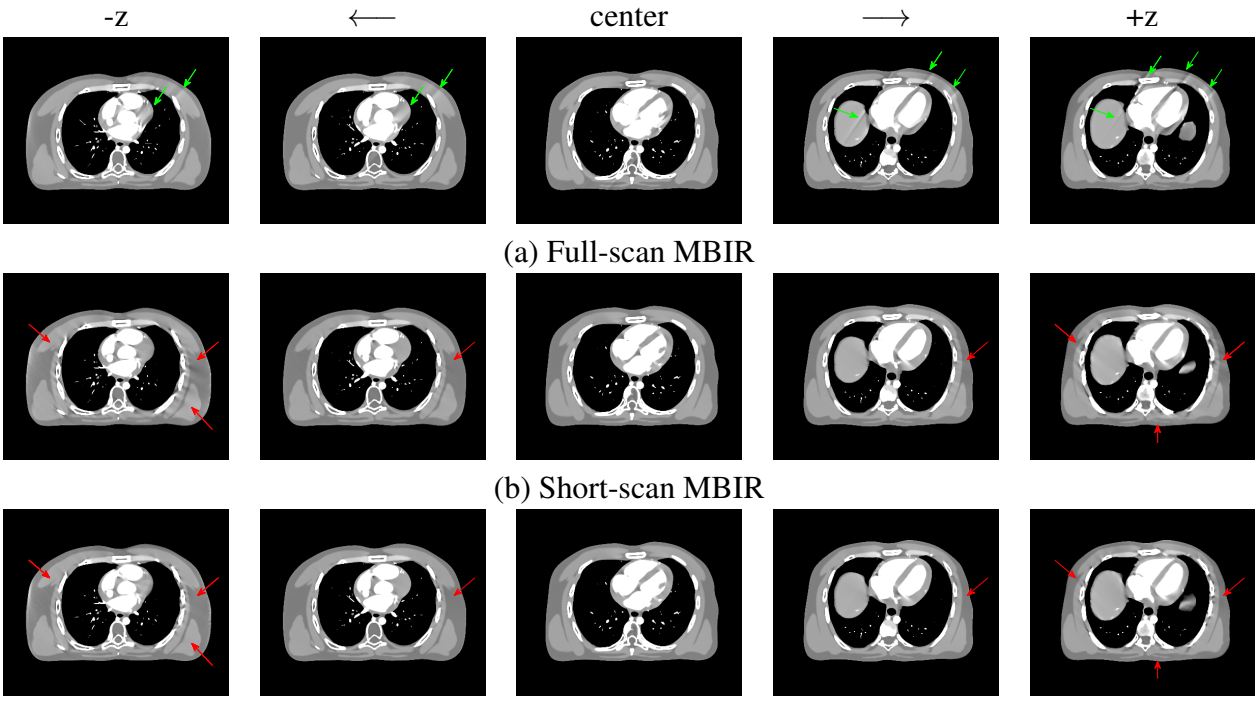


Figure 5.7: View ratio (5.8) values between the proposed method and short-scan MBIR within the football region. From left to right, each column corresponds to 5th, 13th, and center slice of ROI (out of 64 slices), respectively. Display range is [1 1.3].

Fig. 5.8 compares reconstructed images obtained by the proposed method to those of other methods at different slice locations. The reconstructed image obtained by the proposed method removes short-scan artifacts while preserving the temporal resolution in the football region, and thus shows improved image quality compared to that of short-scan MBIR. Notice that the slices on the far end have much less artifacts. However, when compared to reconstructed images from full-scan MBIR, some residual artifacts are still noticeable on end slices. Tuning the design parameter p changes the image quality of the reconstructed image, but does not remove the entire short-scan artifacts while preserving the temporal resolution. This suboptimal results come from the inherent limitation of the statistical weighting modification. Since we only control the contribution of each ray, which affects multiple voxels at the same time, we only have limited control over sampling of each voxel. More sophisticated design for (5.6) may improve the results, but the trade-off between the temporal resolution and the short-scan artifact cannot be avoided with the proposed statistical weighting modification approach.



(a) Full-scan MBIR

(b) Short-scan MBIR

(c) Proposed method (5.6) with $p = 5$

Figure 5.8: Comparison of reconstructed images from different reconstruction methods. (a) Full-scan MBIR (b) Short-scan MBIR (c) Proposed method (5.6) with $p = 5$. From left to right, each column corresponds to 1st, 5th, 32nd, 56th, and 64th slice, respectively (out of 64 slices). Slices from 5th to 56th are used for usual cardiac imaging displays. Green and red arrows indicate motion artifacts and short-scan artifacts, respectively. Display window is [850 1150] (HU).

As a next step, we evaluated the method on the XCAT phantom with wider cone angle, i.e. 256

detector rows. Proposed Proposed method (5.6) method is compared to other methods in Fig. 5.9. The reconstructed image obtained by the proposed Proposed method (5.6) remove the short-scan artifacts while preserving the temporal resolution in the football region, and thus shows improved image quality compared to that of short-scan MBIR. However, when compared to reconstructed images from full-scan MBIR, some residual artifacts are still noticeable on end slices. Tuning the design parameter p changes the image quality of the reconstructed image, but does not remove the entire short-scan artifacts while preserving the temporal resolution. This suboptimal results come from the inherent limitation of the statistical weighting modification. Since we only control the contribution of each ray, which affects multiple voxels at the same time, we only have limited control over sampling of each voxel. More sophisticated design for (5.6) may improve the results, but the trade-off between the temporal resolution and the short-scan artifact cannot be avoided with the proposed statistical weighting modification approach.

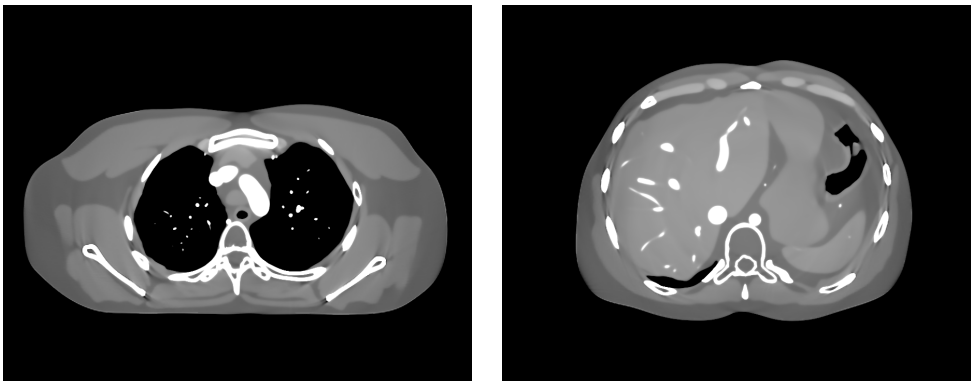
5.1.4 Conclusions for statistical weighting modification

We presented a statistical weighting modification approach to reduce short-scan artifacts while maintaining temporal resolution in the target region. The proposed statistical weighting modification showed improvements compared to short-scan MBIR, especially on end slices. Future work will address other possible designs to further remove residual artifacts. The residual artifact is mostly affected by the targeted region (5.3). Instead of selecting the targeted region based on the sampling, *e.g.*, football region, we can also directly determine the voxels that are mostly affected by the temporal resolution. Having additional measurements enables multi-phase reconstruction. Comparing reconstructed images from adjacent phases could help identifying such voxels, which may be used for more effective statistical weighting modification designs.

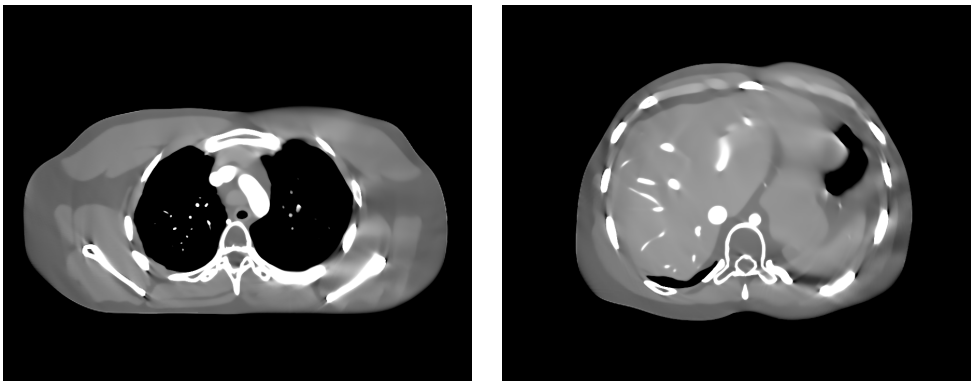
5.2 Introducing additional regularization term

Previously, we have investigated methods to reduce sampling induced artifacts, such as short-scan artifacts, by designing regularizers and modifying statistical weightings. Even though we have investigated various aspects of the regularizer, it was very hard to reduce the artifacts. On the other hand, the statistical weighting modification approach successfully removed some of the artifacts, and this suggests that use of additional information is inevitable to completely remove such artifacts.

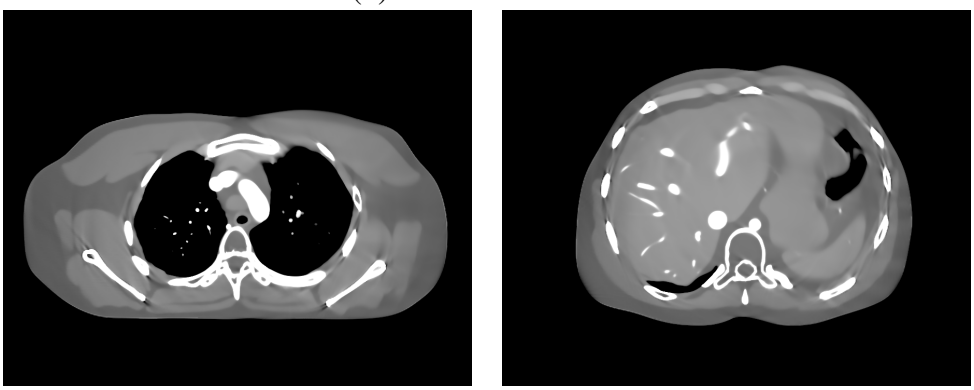
In the MBIR framework, one intuitive approach for the short-scan artifact reduction problem is to use prior information obtained from some image that does not suffer from short-scan artifacts. Such prior images, however, may have other issues such as poor temporal resolution or noise properties. Consider the following PWLS formulation with an additional prior-dependent



(a) Full-scan MBIR



(b) Short-scan MBIR



(c) Proposed method (5.6) with $p = 5$

Figure 5.9: Comparison of reconstructed images from different reconstruction methods. (a) Full-scan MBIR (b) Short-scan MBIR (c) Proposed method (5.6) with $p = 5$. Each column corresponds to 1st and last slice of ROI, respectively. Display window is [800 1200] (HU).

regularizer

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \frac{1}{2} \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_{\boldsymbol{W}}^2 + \mathsf{R}(\boldsymbol{x}) + \beta_p \|\boldsymbol{L}(\boldsymbol{x} - \boldsymbol{x}_{\text{prior}})\|_{\boldsymbol{D}}^2 , \tag{5.9}$$

where β_p is a regularization parameter for the additional prior term, $\boldsymbol{x}_{\text{prior}}$ denotes a prior image, $\boldsymbol{L}$ is a filtering operator, and $\boldsymbol{D}$ selects (or weights) the voxels to be affected by the prior. It is

apparent that the quality of the solution highly depends on that of the prior image. Therefore, the major challenge of this method is to select an appropriate prior image and efficiently exploit its information.

One possible candidate for a prior image is an initial reconstruction using all available measurements. When full-scan measurement is available, we can either use full-scan FBP or full-scan MBIR to obtain images with reduced sampling-related artifacts. Full-scan MBIR would be a better choice in terms of image quality due to its superior noise characteristics, especially for low-dose scans. However, the computation cost issue introduced in Chapter 1.2.3 would becomes more severe since we need to minimize two different cost functions sequentially. On the other hand, full-scan FBP may be readily available and does not require much computational expense compared to full-scan MBIR. However, a FBP prior image may cause degradations of the solution, such as compromised noise property. Therefore, a sophisticated filtering operation is required to extract only the desirable features from the FBP prior.

For some cases, available measurements may not be enough to remove short-scan artifacts from the reconstructed image. To prevent such case, for cardiac scans, we can use a new acquisition protocol that acquires additional measurements to what current short-scan protocol collects, but at lower tube current. Such additional measurements will ensure that we have enough data to obtain an initial reconstruction with desirable image quality without significantly increasing patient dose. Fig. 5.10 shows an example of such acquisition protocols. This new acquisition protocol can also be combined with the statistical weighting modification in previous section. In this study, we assume that the acquired measurements are enough to reconstruct a prior image with reasonable quality, thus we do not use the new protocol.

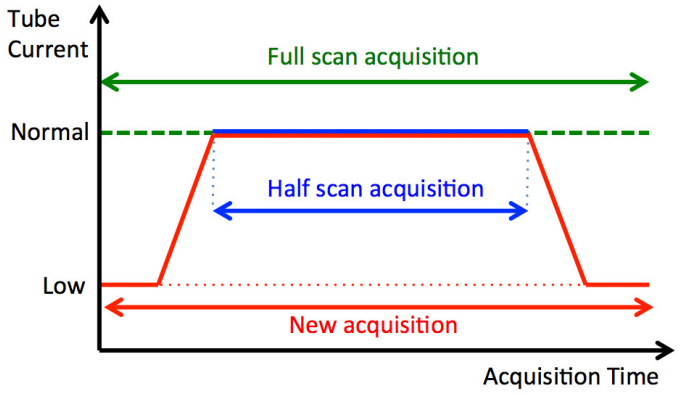


Figure 5.10: Example of a new acquisition protocol

5.2.1 Prior regularization designs

We first define the diagonal weighting $\boldsymbol{D}$ that selects the voxels to be affected by the prior image. Since the prior image may have either poor noise or temporal resolution properties, it is undesirable for sufficiently sampled voxels to be affected by this prior. Unsophisticated designs for the weighting $\boldsymbol{D}$ will result in additional artifacts, degraded temporal resolution and noise properties, etc..

We want to design the weighting based on the sampling condition at each voxel, thus we define a new metric to measure the sampling condition at each voxel location as follows

$$\alpha_j = \frac{\sum_{i=1}^{n_{\mathrm{p}}} a_{ij}}{\sum_{i=1}^{N_{\mathrm{g}}} g_{ij}}, \tag{5.10}$$

where we define $\boldsymbol{G} = \{g_{ij}\}$ to be an extended geometry of $\boldsymbol{A} = \{a_{ij}\}$ with enough number of detector rows to contain the entire reconstruction volume in its football region. For well-sampled region, the sampling metric α_j will have a value close to 1. We also define a sampling mask as follows

$$m_j = 1 - \alpha_j. \tag{5.11}$$

Fig. 5.11 shows an example of the sampling mask at an end slice obtained from a 64-slice CT scanner.

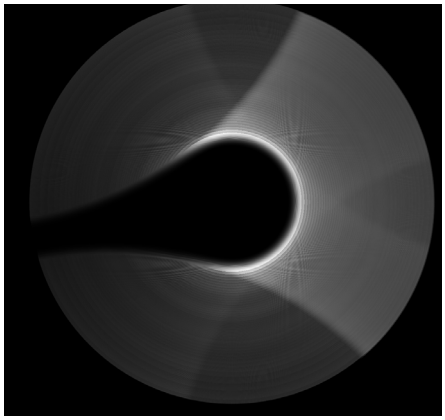


Figure 5.11: A sampling mask ($m_j = 1 - \alpha_j$) at an end slice. A 64-slice CT scanner was used. The color scale goes from 0 (black) to 1 (white).

We propose the weighting $\boldsymbol{D}$ based on the sampling mask (5.11):

$$\boldsymbol{D} = \begin{cases} (m_j)^p & \text{if } j \text{ is in ROI} \\ 0 & \text{otherwise,} \end{cases} \tag{5.12}$$

where p is a parameter that controls the strength of the diagonal weighting.

There are many possible choices for the filtering operator $\boldsymbol{L}$. We first investigate the case without any filtering, *i.e.*, $\boldsymbol{L} = \boldsymbol{I}$. Since the prior image may contain unacceptable noise, no filtering may lead to noisy images, especially at the selected region. We want to investigate how much noise in the prior can be tolerated when no filtering is applied.

Using a low-pass filtering for $\boldsymbol{L}$ seems reasonable, since we only want to remove short-scan artifacts that have mostly low-frequency contents. Properly designed filter will remove noise and unnecessary structures in the prior image and provide useful low-frequency structural information. Since we want to apply the prior regularization to insufficiently sampled region, which is mostly located on end slices, we apply a stack of 2D filtering instead of a 3D filtering. We use the following definition for the filtering operator $\boldsymbol{L}$

$$\boldsymbol{L} = \begin{bmatrix} \vdots \\ \boldsymbol{M}_i' \boldsymbol{Q}' \boldsymbol{L}_i \boldsymbol{Q} \boldsymbol{M}_i \\ \vdots \end{bmatrix}, \tag{5.13}$$

where $\boldsymbol{L}_i$ is a 2D Fourier domain filter, $\boldsymbol{M}_i$ is a mask that selects ith slice from the entire volume, and $\boldsymbol{Q}$ is a 2D DFT matrix. For 2D low-pass filter $\boldsymbol{L}_i$, we investigate the following two simple choices Fig. 5.12.

- Rectangular filter
$$L_i(u,v) = \begin{cases} 1 & \text{if } r(u,v) < F_c \\ 0 & \text{otherwise,} \end{cases}$$
- Gaussian filter
$$L_i(u,v) = \exp\left(-\frac{r^2}{2F_c^2}\right),$$
where $u = 0, 1, ..., N_x - 1$, $v = 0, 1, ..., N_y - 1$, F_c is a normalized cutoff frequency, and
$$r(u,v) = \sqrt{((u - N_x/2)/N_x)^2 + ((v - N_y/2)/N_y)}.$$

Rectangular filter was designed with a smooth transition between the stop and pass band to avoid ringing artifacts.

5.2.2 Results for additional prior regularization

Since the diagonal weighting $\boldsymbol{D}$ (5.12) can preserve temporal resolution of the well-sampled region, we only focus on reduction of short-scan artifacts in the undersampled region. For this

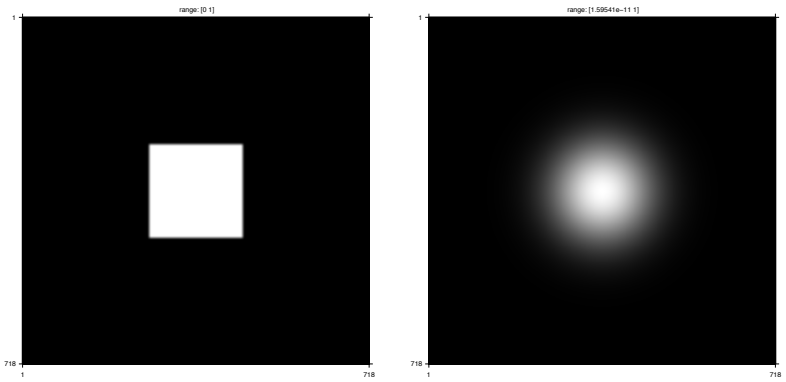


Figure 5.12: 2D Low-pass filters: Rectangular (left) and Gaussian (right) with $F_c = 0.2$.

purpose, we used a static XCAT phantom for following experiments. All the reconstructed images were obtained by using a "order-subsets (OS) + momentum" method based on Nesterov's fast gradient method [75] with 12 subsets and 30 iterations.

5.2.2.1 Importance of the ROI constraint

It is important to properly weight the prior image such that the prior only affect the voxels within ROI (see (5.12)). Reconstructed images outside ROI contains artifacts or may not be consistent with the system model, especially when FBP based prior is being used. Enforcing prior regularization on such regions can lead to additional artifacts. We illustrate this with the following example with the XCAT phantom. A 256-slice scanner was assumed and no noise was added to the measurements. Full-scan FBP reconstruction was used as the prior image. Fig. 5.13 shows the error sinogram, $\boldsymbol{y} - \boldsymbol{A}\hat{\boldsymbol{x}}$, of reconstructed images obtained with and without the ROI constraint. Severe mismatch in error sinogram can be observed when the ROI constraint is not enforced. This suggests that the prior image is not consistent with the system model, thus strong enforcement of the prior will results in artifacts. Fig. 5.14 and Fig. 5.15 compare the reconstructed images obtained with and without the ROI constraint. Reconstructed images were obtained from the design (5.13) with $\boldsymbol{L} = \boldsymbol{I}$ and $\beta_p = 10^{10}$. Images obtained without the constraint are corrupted with artifacts. Artifacts mostly appear around the transition region of the diagonal weighting $\boldsymbol{D}$.

Fig. 5.16 illustrates that use of the rectangular filter can cause ringing artifacts within the reconstructed image, especially around high contrast structures. The transition in our rectangular filter was not smooth enough to avoid the ringing artifact. More improved designs with better transition and parameters may reduce such artifacts, but designing a sophisticated filter is not our goal in this study. From now on, we focus on designs with Gaussian filters.

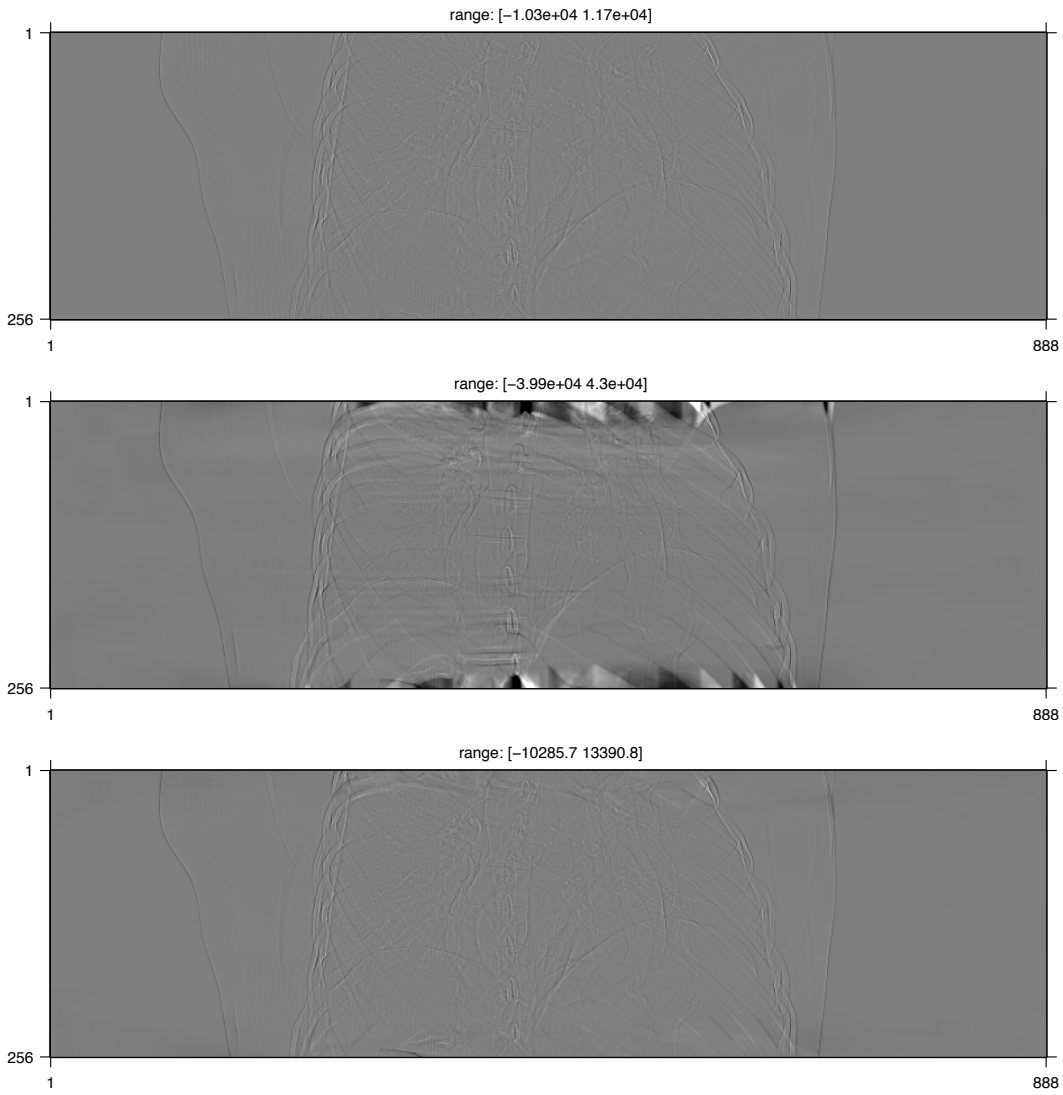


Figure 5.13: Comparison of error sinograms of the following reconstructed images: Short-scan MBIR (top), Design (5.13) without ROI constraint for the diagonal weighting $\boldsymbol{D}$ (middle), Design (5.13) with ROI constraint for $\boldsymbol{D}$ (bottom).

5.2.2.2 Prior with moderate noise: a regular-dose case

We simulated a regular-dose case, which has a tube current of 550mAs. A 256-slice scanner was assumed and the XCAT phantom was used as a static object. Realistic noise were simulated with CATSIM [33]. Full-scan FBP image is used as the prior image.

Fig. 5.17 and Fig. 5.18 present reconstructed images obtained from various methods. Regularization parameter for the prior regularization, β_p, was selected empirically for the purpose of illustration. Full-scan FBP image used as the prior is noisier than full-scan or short-scan MBIR images. The reconstructed image obtained from the proposed design with no filtering had much

(a) without ROI constraint

(b) with ROI constraint

Figure 5.14: Comparison between the reconstructed images obtained using the design (5.13) (a) without ROI constraint (b) with ROI constraint. From left to right, each column corresponds to 1st, 17th, center, 239th, last slice of ROI, respectively. Display window is [800 1200] (HU).

(a) without ROI constraint

(b) with ROI constraint

Figure 5.15: Comparison between the reconstructed images obtained using the design (5.13) (a) without ROI constraint (b) with ROI constraint. Sagittal and coronal views. Display window is [800 1200] (HU).

less short-scan artifacts, but its end slices are noisy compared to its middle slices. In this case, the noise characteristics of the prior image were propagated to the reconstructed image due to the absence of filtering. This was somewhat expected from the beginning, and the regularizer $\mathsf{R}(\boldsymbol{x})$ was unable to remove the propagated noise. The main reason for this situation is that the regularization strength of $\mathsf{R}(\boldsymbol{x})$ at those locations considered only the data fitting term, and not the

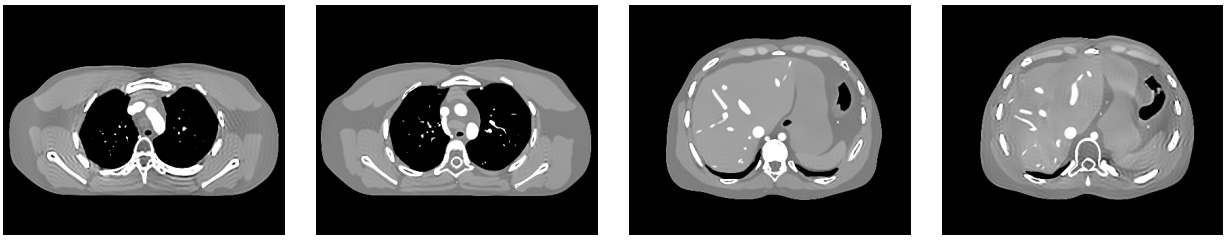


Figure 5.16: Reconstructed images using the proposed design with a rectangular filter ($F_c = 1/5$ and $\beta_p = 10^{10}$). Each column corresponds to 1st, 17th, 239th, last slice of ROI, respectively. Display window is [800 1200] (HU).

additional prior regularization term. If we want to remove such noise, we need to properly modify the regularization strength of the regularizer $\mathsf{R}(\boldsymbol{x})$ by considering the effect of the prior term. This is an open question, and we leave it as a future work. The reconstructed image obtained from the proposed method with a Gaussian filter, on the other hand, had much better noise property at end slices, in addition to reduced short-scan artifacts. Use of low-pass filtering stopped noise and other high-frequency artifacts from propagating into the reconstructed image. However, when compared to full-scan MBIR image, the reconstructed image still shows residual artifacts and is slightly noisier at end slices.

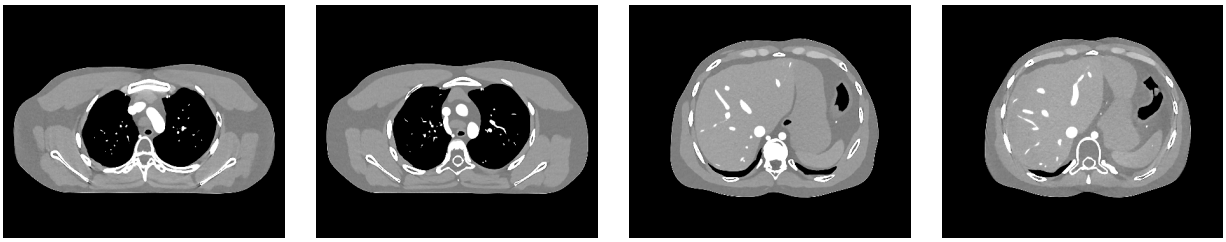
So far, we fixed the cutoff frequency to $F_c = 1/5$. When the cutoff frequency is too high, then the noise and undesirable high-frequency structures can propagate into the reconstructed image. We investigated different values of the cutoff frequency to see its effect on the image. Fig. 5.19 compares reconstructed images obtained from different cutoff frequencies. When the cutoff frequency becomes too small, the reconstructed image loses some of its details and suffers from other artifacts. Empirically, $F_c = 1/10$ provided well-balanced results.

5.2.2.3 Prior with severe noise: a low-dose case

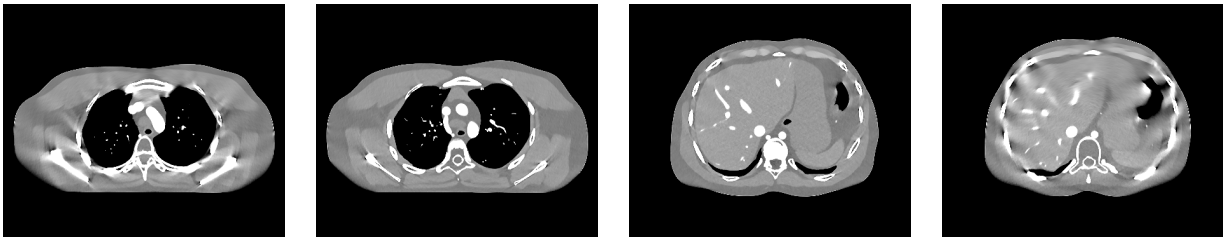
To test a case where the prior image is too corrupted with noise and artifacts to provide accurate information, we simulated a low-dose case with 50mAs tube current. All other settings are the same as the regular-dose case.

Fig. 5.20 and Fig. 5.21 show reconstructed images from various methods. For this case, the prior image, Full-scan adaptive statistical iterative reconstruction (ASiR$^{\text{TM}}$)[5] is noisy and have severe artifacts. For the locations where the prior image is good enough to provide useful structural information, the reconstructed image from the proposed method shows reduced artifacts and desired noise characteristics. However, severe artifacts of the prior image were also propagated at

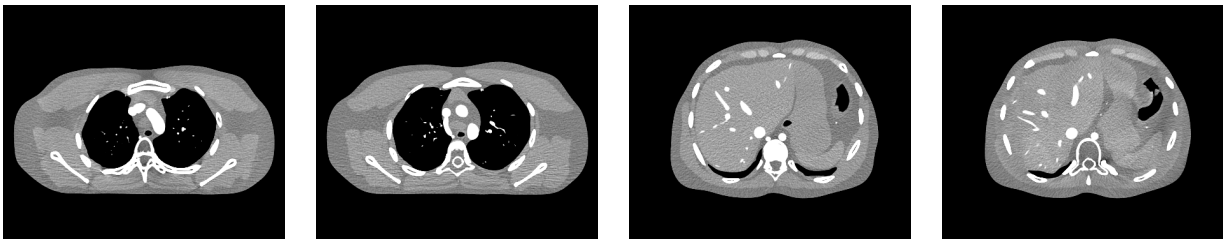
[5]ASIR image was provided by GE Healthcare. However, the image is not from the full product ASIR meaning that it may be missing some pre- or post- processing that could affect the reconstructed image quality.



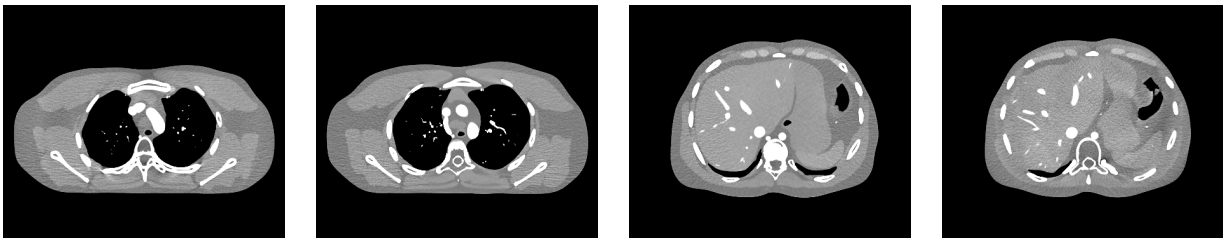
(a) Full-scan MBIR



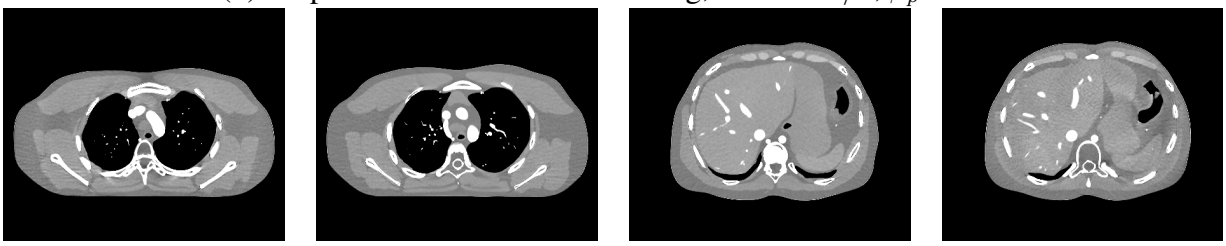
(b) Short-scan MBIR



(c) Full-scan FBP



(d) Proposed method with no filtering, cutoff = $1/5$, $\beta_p = 10^6$



(e) Proposed method with a Gaussian filter, cutoff = $1/5$, $\beta_p = 10^6$

Figure 5.17: Comparison between the reconstructed images obtained from various methods (a) Full-scan MBIR (b) Short-scan MBIR (c) Full-scan FBP (d) Proposed method with no filtering (e) Proposed method with a Gaussian filter. From left to right, each column corresponds to 1st, 17th, 239th, last slice of ROI, respectively. Display window is [800 1200] (HU).

some locations. This experiment showed the limitation of the prior based regularization. The quality of the reconstructed image depends heavily on that of the prior image, and imperfect design of the regularization may lead to additional artifacts.

In Fig. 5.22 and Fig. 5.23, we present reconstructed images obtained by using the statistical weighting modification approach proposed in the previous section. The statistical weighting modification greatly reduced short-scan artifacts. However, as pointed out in the previous section, due to the limitation of the statistical weighting modification, there exists residual artifacts in the image. To reduce the artifact further, we combined the prior regularization approach with the statistical weighting approach. Fig. 5.22 and Fig. 5.23 also show the reconstructed image from the combined method. We selected the regularization parameter for the prior regularization term to be small, $\beta_p = 10^3$. Combined method slightly reduced residual artifacts.

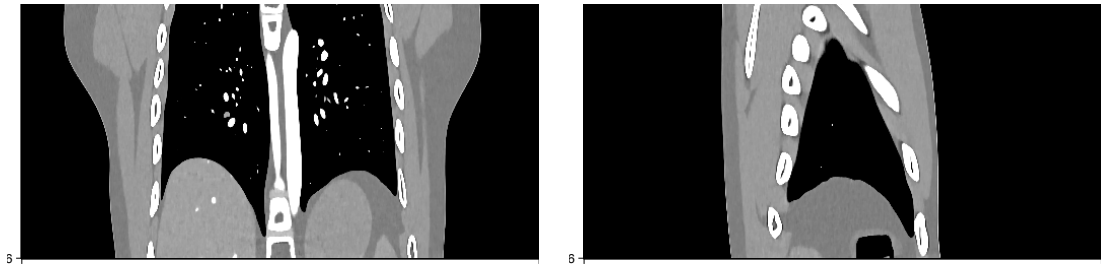
5.2.3 Discussion for additional prior regularization

Adding an additional regularization based on the prior image showed promising results. By designing the diagonal weighting or tuning the parameters, we can only use the necessary information for the reconstruction. However, it is hard to tune all the parameters optimally. How to determine each of these parameters is an open question. In addition, the quality of the prior image greatly affects the quality of the reconstructed image. We used FBP based prior images for the experiments, but we can use full-scan MBIR instead for better results. However, this requires much more computation, since we are reconstructing two separate MBIR images.

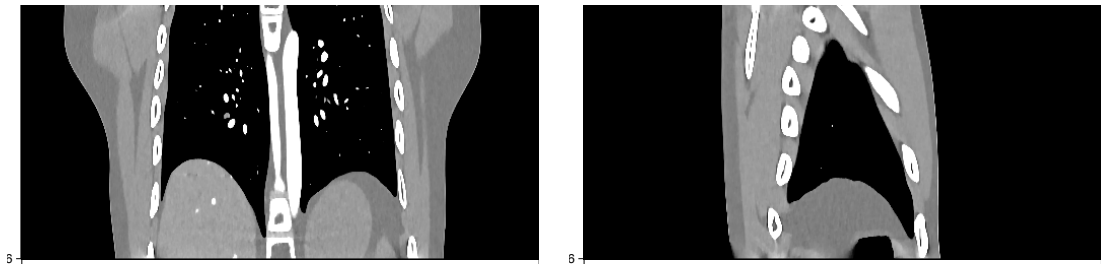
We also showed that the proposed regularization method can be combined with the statistical weighting modification method to further reduce the short-scan artifacts.

5.3 Summary of the chapter

In this chapter, we investigated methods that reduce short-scan artifacts by exploiting all available information. We used additional information other than short-scan measurements to reconstruct images with reduced artifacts. The main issues were how to effectively use such additional information and to maintain desirable properties such as temporal resolution. We proposed two different methods to achieve the goal: statistical weighting modification and additional prior regularization. Both methods greatly reduce short-scan artifacts in the reconstruct image. However, they both suffer from certain trade-offs. The main question for these methods is the selection of the parameters that control these trade-offs. How to choose such parameters is an open question. To systematically choose the parameters, we need a metric that can measure the amount of short-scan artifact in the image. However, it is not trivial to design the metric since the artifacts appear in various forms. We leave the metric design as a future work.



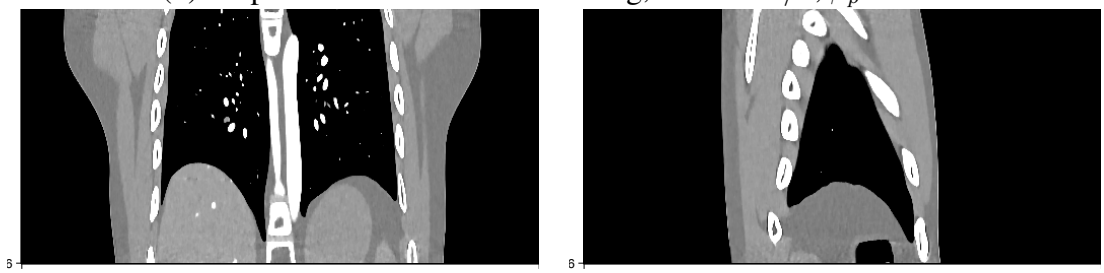
(a) Full-scan MBIR



(b) Short-scan MBIR

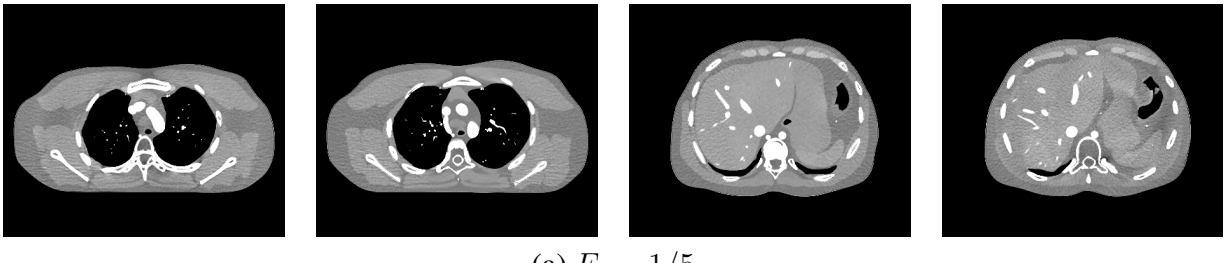
(c) Full-scan FBP

(d) Proposed method with no filtering, cutoff = $1/5$, $\beta_p = 10^6$

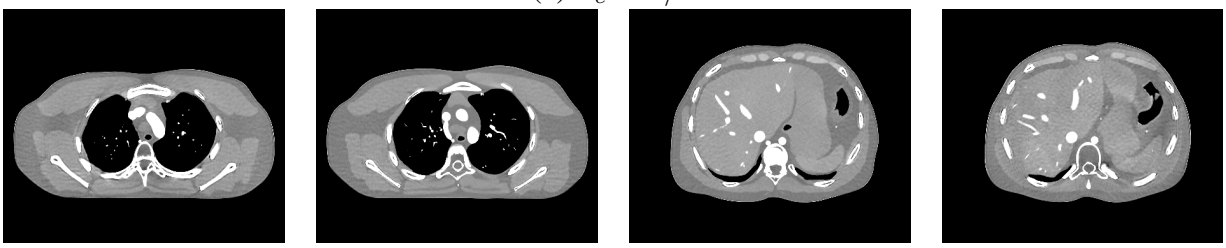


(e) Proposed method with a Gaussian filter, cutoff = $1/5$, $\beta_p = 10^6$

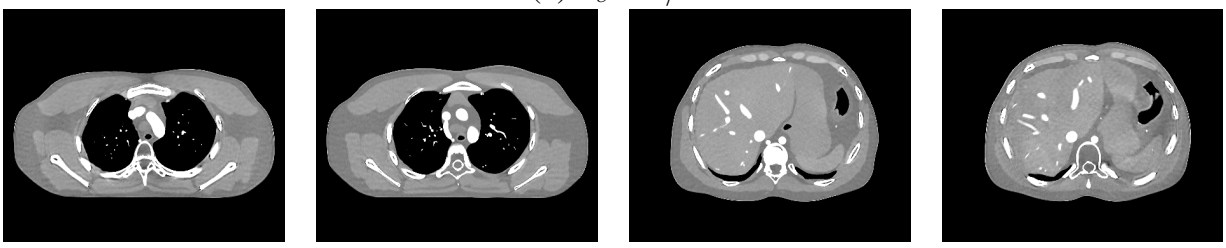
Figure 5.18: Comparison between the reconstructed images obtained from various methods (a) FS-MBIR (b) HS-MBIR (c) FSHS-FBP (d) Proposed method with no filtering (e) Proposed method with a Gaussian filter. Sagittal and coronal views. Display window is [800 1200] (HU).



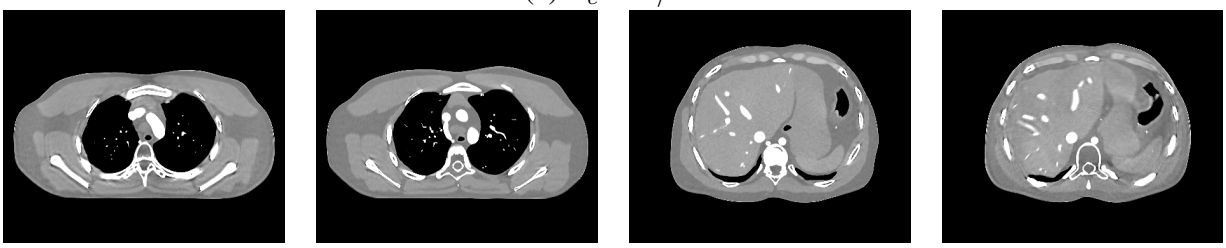
(a) $F_c = 1/5$



(b) $F_c = 1/8$

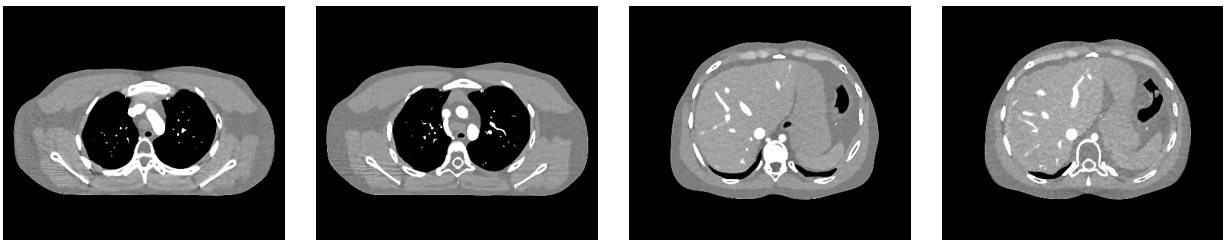


(c) $F_c = 1/10$

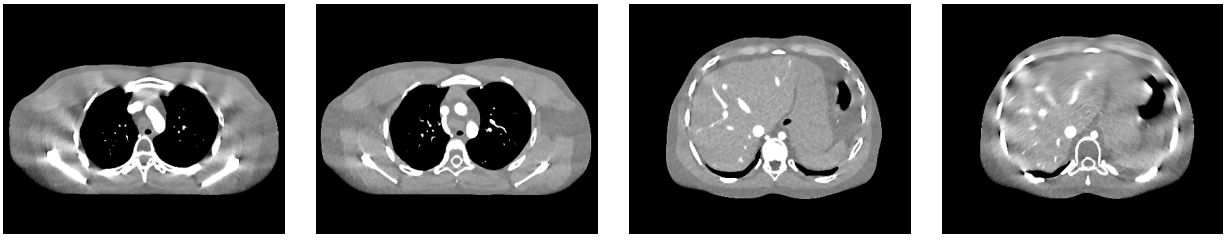


(d) $F_c = 1/15$

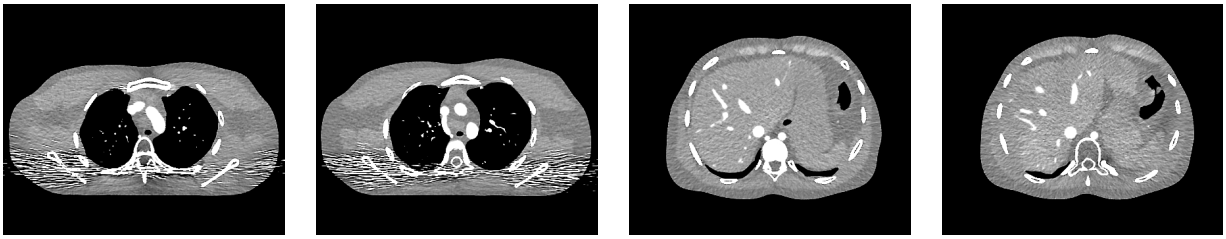
Figure 5.19: Reconstructed images obtained from different cutoff frequencies (a) $F_c = 1/8$ (b) $F_c = 1/10$ (c) $F_c = 1/15$. Each column corresponds to 1st, 17th, 239th, last slice of ROI, respectively. Display window is [800 1200] (HU).



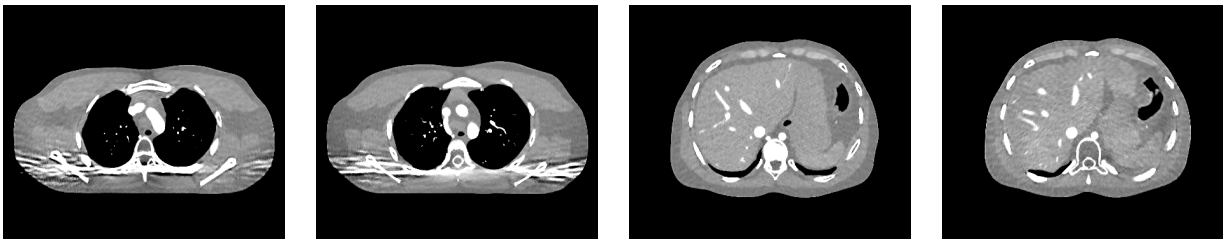
(a) Full-scan MBIR



(b) Short-scan MBIR



(c) Full-scan ASiR

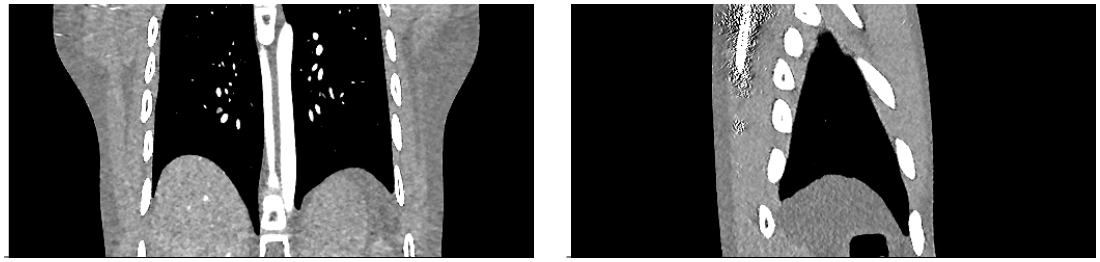


(d) Proposed method with Gaussian filtering, cutoff = $1/10$, $\beta_p = 10^6$

Figure 5.20: Comparison between the reconstructed images obtained from various methods (a) Full-scan MBIR (b) Short-scan MBIR (c) Full-scan ASIR (d) Proposed method with a Gaussian filter. From left to right, each column corresponds to 1st, 17th, 239th, last slice of ROI, respectively. Display window is [800 1200] (HU).

(a) Full-scan MBIR

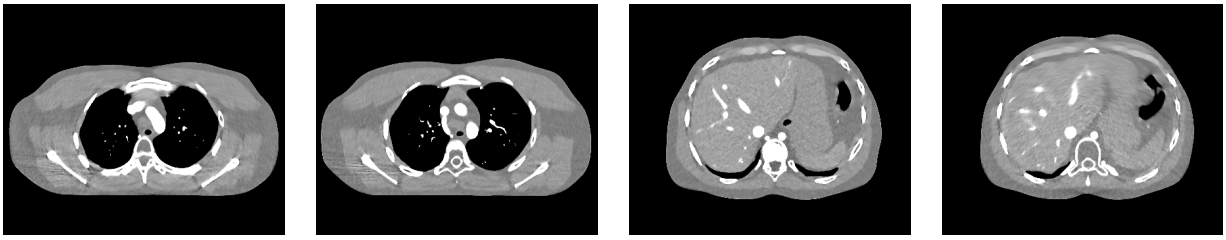
(b) Short-scan MBIR



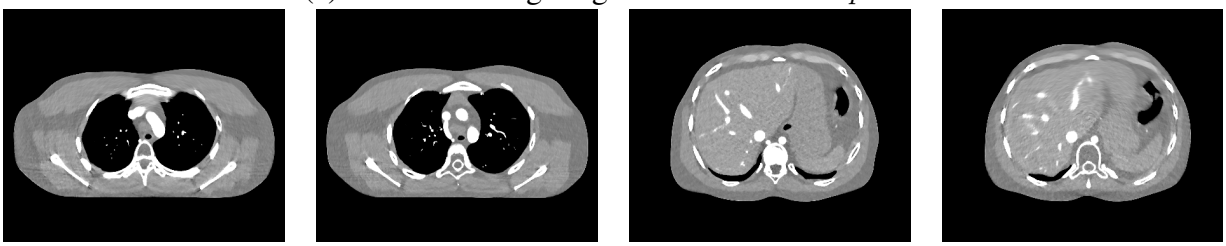
(c) Full-scan ASiR

(d) Proposed method with Gaussian filtering, cutoff = $1/10$, $\beta_p = 10^6$

Figure 5.21: Comparison between the reconstructed images obtained from various methods (a) FS-MBIR (b) HS-MBIR (c) FSHS-FBP (d) Proposed method with no filtering (e) Proposed method with a Gaussian filter. Sagittal and coronal views. Display window is [800 1200] (HU).



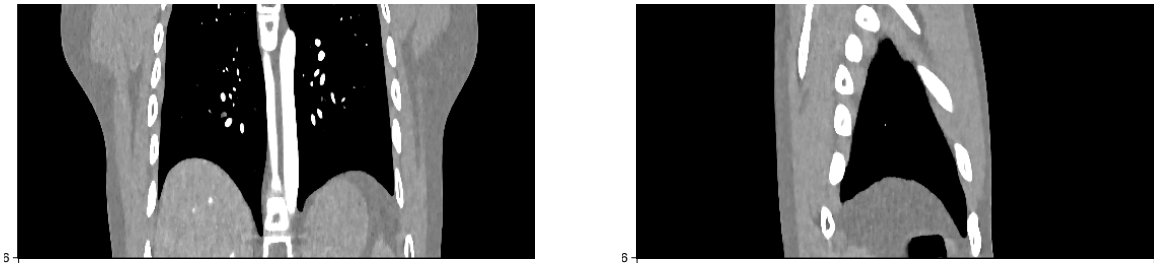
(a) Statistical weighting modification with $p = 5$



(b) Statistical weighting modification + the prior regularization method with Gaussian filtering

Figure 5.22: Reconstructed images from (a) statistical weighting modification (b) statistical weighting modification ($p = 5$) combined with the prior regularization method with Gaussian filtering ($F_c = 1/10$ and $\beta_p = 10^3$). Each column corresponds to 1st, 17th, 239th, last slice of ROI, respectively. Display window is [800 1200] (HU).

(a) Statistical weighting modification



(b) Statistical weighting modification + the prior regularization method with Gaussian filtering

Figure 5.23: Reconstructed images from (a) statistical weighting modification (b) statistical weighting modification ($p = 5$) combined with the prior regularization method with Gaussian filtering ($F_c = 1/10$ and $\beta_p = 10^3$). Sagittal and coronal views. Display window is [800 1200] (HU).

CHAPTER VI

Accelerated statistical image reconstruction methods

In CT imaging. the complex nature of the cost function that comes from the complicated and large-scaled system matrix and a generally non-quadratic regularizer necessitate the need of an iterative algorithm for the minimization process. As described in Section 1.2.3, the computation time is one of the great challenges for the practical utilization of the statistical image reconstruction method. Choice of an algorithm and its properties such as monotonicity and convergence speed are directly related to the computational expense, so many algorithms with advanced properties have been proposed in many literatures [8, 75, 105, 140]. This chapter presents methods to accelerate the optimization process for both cases: a standard image reconstruction without motion model and a motion-compensated image reconstruction.

6.1 Accelerating ordered-subsets type image reconstruction using double surrogates

Ordered-subsets (*OS*) algorithms, also known as block-iterative or incremental gradient methods, are popular in the field of statistical image reconstruction due to their significant acceleration in initial iterations. The basic idea of OS methods is to group the measurement data into an ordered sequence of subsets or blocks and utilize only one subset of the data for each update instead of using the entire measurements.

Despite their success in speeding up the convergence, ordinary OS algorithms are not convergent in general but rather approach a suboptimal limit-cycle. Several methods were proposed to make OS algorithms globally convergent [5, 76, 117]. Most of these methods introduces relaxiation parameters, which need to be appropriately scheduled to ensure global convergence. Finding good relaxation parameters usually involves trial-and-error or some experimentation [5]. There also exist convergent algorithms that does not use relaxation parameters such as incremental EM

This chapter is based on [17, 18, 24, 25]

algorithms [86] and incremental aggregated gradient (IAG) method [12]. Incremental optimization transfer (IOT) is generalized version of the incremental EM algorithms, and it ensures global convergence without requiring relaxation parameters [6].

Unregularized OS reconstruction methods are used routinely for PET and SPECT scans [59]. For regularized image reconstruction problems, conventional OS algorithms calculate the gradient of the regularizer for every subset update [1]. For large problems with large number of subsets, such as cone-beam or helical CT image reconstruction, calculating the gradient for each subset update can be very computationally expensive. To mitigate this issue, some investigators use unregularized iterations followed by a denoising operation after updating all subsets [135]. Although such methods save computation, their convergence properties are uncertain, and since they may not be minimizing any particular cost function it becomes more difficult to design regularization parameters. Furthermore, it is known that inserting filtering steps into unregularized algorithms can lead to undesirable spatial resolution properties [84]. Our goal here is to reduce the computational cost without inducing such problems.

In this section, we propose new OS-type algorithms that are derived from optimization transfer principles and that allow us to compute the gradient of the regularizer less frequently, and thus reduces the computational expense when many subsets are used. Our derivation leads to a correction term that accounts for the fact that the regularizer gradient is updated less frequent than every subiteration. The proposed method reduces computational cost with little impact on the convergence rate leading to an overall acceleration.

Recently, OS algorithms were further accelerated by using spatially non-uniform updates [74] or incorporating momentum method based on Nesterov's fast gradient method [73, 75]. These new OS algorithms converge much faster than the conventional OS, even with small number of subsets. We extend the proposed algorithm to OS with momentum method, and show that, even when small number of subsets are being used, our method can still save computation time.

We apply the proposed algorithm to penalized weighted least squares (*PWLS*) image reconstruction for cone-beam X-ray computed tomography (*CT*). Simulations and phantom experiments show that the proposed method reconstructed images with compatible image quality within reduced computation time. The method easily could be adapted for other statistical models.

6.1.1 Ordered-subsets (OS) and incremental optimization transfer (IOT) algorithms

6.1.1.1 Ordered-subsets algorithms

For most penalized-likelihood image reconstruction problems, the objective function and its gradient can be rewritten as the following partially separable forms [88]:

$$\Psi(\boldsymbol{x}) = \sum_{m=1}^{M} \Psi_m(\boldsymbol{x}), \quad \nabla_{\boldsymbol{x}} \Psi(\boldsymbol{x}) = \sum_{m=1}^{M} \nabla_{\boldsymbol{x}} \Psi_m(\boldsymbol{x}), \tag{6.1}$$

where typically $\Psi_m(\boldsymbol{x})$ corresponds to a subset of the projection data. Most iterative algorithms use the gradient of the objective function to obtain a minimizer, and many can be written as the form

$$\boldsymbol{x}^{(n+1)} = \boldsymbol{x}^{(n)} - \alpha_n \boldsymbol{D}_0(\boldsymbol{x}^{(n)}) \nabla_{\boldsymbol{x}} \Psi(\boldsymbol{x}^{(n)}), \tag{6.2}$$

where $\alpha_n > 0$ is a relaxation parameter and $\boldsymbol{D}_0$ is a scaling matrix, which is typically diagonal. For many algorithms, the empirical findings have suggested that in early iterations, one can replace the gradient of the entire objective function with that of only a part of the objective function and still have $\boldsymbol{x}^{(n+1)}$ be better than $\boldsymbol{x}^{(n)}$ but with less computational cost. Such methods are called incremental gradient methods in the optimization literature [10, 72]. The term ordered-subsets has been used in tomography fields, because only a subset of projection views are used in the update steps and the ordering of the views is important [1, 59, 71]. For $\boldsymbol{x}^{(n)}$ far from the minimizer, $\hat{\boldsymbol{x}}$, if we select the subsets to be "balanced" in some appropriate sense then the following conditions hold:

$$\nabla_{\boldsymbol{x}} \Psi_1(\boldsymbol{x}) \cong \nabla_{\boldsymbol{x}} \Psi_2(\boldsymbol{x}) \cong \cdots \cong \nabla_{\boldsymbol{x}} \Psi_M(\boldsymbol{x}),$$

or equivalently,

$$\nabla_{\boldsymbol{x}} \Psi(\boldsymbol{x}) \approx M \nabla_{\boldsymbol{x}} \Psi_m(\boldsymbol{x}), \quad \forall m. \tag{6.3}$$

Thus, instead of (6.2), a typical OS algorithm has the form shown in Table 6.1.

We refer to each update in Table 6.1 as the mth subset update, or mth sub-iteration of the nth iteration. One complete iteration is performed by cycling through all the subsets indexed by m so that all data is utilized. In tomography problems, the subsets are selected so that projections within each subset correspond to angularly downsampled projections. It was suggested that the ordering of the subsets that makes projections corresponding to one subset as "orthogonal" as possible to previously used projections is preferable [49, 54]. Despite their success in speeding up the initial convergence, ordinary OS algorithms are not convergent in general but rather approach a suboptimal limit-cycle without relaxation, i.e., when $\alpha_n = \alpha$. To address this issue, several families of convergent OS type algorithms have been proposed [12, 34, 76, 86], although those

Table 6.1: Ordered-subsets algorithm

Initialize $\boldsymbol{x}^{(0)}$

for $n = 0, 1, \cdots,$

$$\boldsymbol{x}^{(n)} := \boldsymbol{x}^{(n,0)}$$

for $m = 1, \cdots, M$

$$\boldsymbol{x}^{(n,m)} = \boldsymbol{x}^{(n,m-1)} - \alpha_n M \boldsymbol{D}(\boldsymbol{x}^{(n,m-1)}) \nabla_{\boldsymbol{x}} \Psi_m\left(\boldsymbol{x}^{(n,m-1)}\right),$$

end

$$\boldsymbol{x}^{(n+1)} := \boldsymbol{x}^{(n,M)}$$

end

modifications tend to slow down convergence. In this section, we focus on the initial convergence characteristics of OS algorithms rather than their final convergence properties.

Consider a PWLS objective function of the form (3.1). The data-fit term can be rewritten as the following separable form:

$$\mathsf{L}(\boldsymbol{x}) = \sum_{m=1}^{M} \mathsf{L}_m(\boldsymbol{x}),$$

where L_m is the data-fit term associated with the mth subset. OS algorithm for PWLS objective function is shown in Table 6.2. For *separable quadratic surrogates* method, we can define the scaling matrix $\boldsymbol{D}$ as follows:

$$\boldsymbol{D}(\boldsymbol{x}) = \left[\boldsymbol{D}_L(\boldsymbol{x}) + \boldsymbol{D}_R(\boldsymbol{x})\right]^{-1}.$$

Typical choices [1] for $\boldsymbol{D}_L$ and $\boldsymbol{D}_R$ are

$$\begin{aligned} \boldsymbol{D}_L &= \operatorname{diag}(\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A}\mathbf{1}), \\ \boldsymbol{D}_R(\boldsymbol{x}) &= \operatorname{diag}\left(|\boldsymbol{C}|' \operatorname{diag}(|\omega(\boldsymbol{C}\boldsymbol{x})|)\, |\boldsymbol{C}|\, \mathbf{1}\right), \end{aligned}$$

where $\omega(t) = \dot{\psi}(t)/t$ is Huber's curvature [58]. Instead of using pre-determined scaling factors, more adaptive selection maybe desirable. For example, OS with spatially non-uniform updates [74] refines these scaling matrices based on the prediction of how much updates does each voxel need. Notice that we need to calculate the regularization gradient, $\nabla_{\boldsymbol{x}} \mathsf{R}\left(\boldsymbol{x}^{(n,m-1)}\right)$, for every sub-iteration, which may cause each update to be very computationally expensive when dealing with large problems with many subsets. Even if we used small number of subsets, if the reconstruction volume is large, the computational cost for updating the regularizer gradient would become burdensome.

Table 6.2: Ordered-subsets algorithm for PWLS problem

Initialize $\boldsymbol{x}^{(0)}$

for $n = 0, 1, \cdots,$

$$\boldsymbol{x}^{(n)} \; := \boldsymbol{x}^{(n,0)}$$

for $m = 1, \cdots, M$

$$\boldsymbol{x}^{(n,m)} = \boldsymbol{x}^{(n,m-1)} - \alpha_n \boldsymbol{D}(\boldsymbol{x}^{(n,m-1)}) \left(M \nabla_{\boldsymbol{x}} \text{Ł}_m(\boldsymbol{x}^{(n,m-1)}) + \nabla_{\boldsymbol{x}} \mathsf{R}\big(\boldsymbol{x}^{(n,m-1)}\big) \right),$$

end

$$\boldsymbol{x}^{(n+1)} \; := \boldsymbol{x}^{(n,M)}$$

end

6.1.1.2 Incremental optimization transfer

Consider the separable form of the objective function again (6.1). Assume furthermore that, for each sub-objective function $\Psi_m(\boldsymbol{x})$, we can define a surrogate function ϕ_m that satisfies the usual majorization conditions [60]:

$$\begin{aligned} \phi_m(\boldsymbol{x}; \boldsymbol{x}) &= \Psi_m(\boldsymbol{x}), \quad \forall \boldsymbol{x} \\ \phi_m(\boldsymbol{x}; \bar{\boldsymbol{x}}) &\geq \Psi_m(\boldsymbol{x}), \quad \forall \boldsymbol{x}, \bar{\boldsymbol{x}}. \end{aligned} \tag{6.4}$$

We define the following "divergence" functions:

$$\boldsymbol{D}_m(\boldsymbol{x} \| \bar{\boldsymbol{x}}) \triangleq \phi_m(\boldsymbol{x}; \bar{\boldsymbol{x}}) - \Psi_m(\boldsymbol{x}), \quad m = 1, \cdots, M. \tag{6.5}$$

By construction, $\boldsymbol{D}_m(\boldsymbol{x} \| \bar{\boldsymbol{x}}) \geq 0$ and $\boldsymbol{D}_m(\boldsymbol{x} \| \boldsymbol{x}) = 0$, and especially:

$$\arg\min_{\bar{\boldsymbol{x}}} \boldsymbol{D}_m(\boldsymbol{x} \| \bar{\boldsymbol{x}}) = \boldsymbol{x}.$$

We define the following augmented objective function using (6.5):

$$\bar{\mathsf{F}}(\boldsymbol{x}; \bar{\boldsymbol{x}}_1, \cdots, \bar{\boldsymbol{x}}_M) = \Psi(\boldsymbol{x}) + \sum_{m=1}^{M} \boldsymbol{D}_m(\boldsymbol{x} \| \bar{\boldsymbol{x}}_m) = \sum_{m=1}^{M} \phi_m(\boldsymbol{x} \| \bar{\boldsymbol{x}}_m). \tag{6.6}$$

By the properties of $\boldsymbol{D}_m(\cdot \| \cdot)$, we obtain the following expression for the estimator $\hat{\boldsymbol{x}}$:

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \Psi(\boldsymbol{x}) = \arg\min_{\boldsymbol{x}} \min_{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_M} \bar{\mathsf{F}}(\boldsymbol{x}; \bar{\boldsymbol{x}}_1, \cdots, \bar{\boldsymbol{x}}_M). \tag{6.7}$$

Therefore the problem of minimizing Ψ is equivalent to minimizing $\bar{\mathsf{F}}$. Given an initial set of estimates $\bar{\boldsymbol{x}}_m^{(0)}$ for $m = 1, \cdots, M$, we minimize $\bar{\mathsf{F}}$ by applying a cyclic coordinate descent approach in which we alternate between updating $\boldsymbol{x}$ and one of the $\bar{\boldsymbol{x}}_m$ vectors, as outlined in Fig. 6.3.

Each update will decrease $\bar{\mathsf{F}}$ monotonically, and by (6.7), if the iterates converge to a minimizer of $\bar{\mathsf{F}}$, then that estimates also minimizes the original objective function Ψ. However, Ψ need not decrease monotonically with the algorithm [15].

One drawback of this method is the requirement of storing at least M images, especially when the image size and M are large.

Table 6.3: Generic incremental optimization transfer (IOT) method

Initialize $\bar{\boldsymbol{x}}_1^{(0)}, \cdots, \bar{\boldsymbol{x}}_M^{(0)}$
for $n = 0, 1, \cdots$
  for $m = 1, \cdots, M$

$$\begin{aligned} \boldsymbol{x}^{new} &= \arg\min_{\boldsymbol{x}} \bar{\mathsf{F}}(\boldsymbol{x}; \bar{\boldsymbol{x}}_1^{(n+1)}, \cdots, \bar{\boldsymbol{x}}_{m-1}^{(n+1)}, \bar{\boldsymbol{x}}_m^{(n)}, \bar{\boldsymbol{x}}_{m+1}^{(n)}, \cdots, \bar{\boldsymbol{x}}_M^{(n)}) \\ \bar{\boldsymbol{x}}_m^{(n+1)} &= \arg\min_{\bar{\boldsymbol{x}}_m} \bar{\mathsf{F}}(\boldsymbol{x}^{new}; \bar{\boldsymbol{x}}_1^{(n+1)}, \cdots, \bar{\boldsymbol{x}}_{m-1}^{(n+1)}, \bar{\boldsymbol{x}}_m, \bar{\boldsymbol{x}}_{m+1}^{(n)}, \cdots, \bar{\boldsymbol{x}}_M^{(n)}) = \boldsymbol{x}^{new} \end{aligned} \tag{6.8}$$

  end

$$\boldsymbol{x}^{(n+1)} := \bar{\boldsymbol{x}}_M^{(n+1)}$$

end

When we use quadratic surrogates for ϕ_m, i.e.

$$\phi_m(\boldsymbol{x}; \bar{\boldsymbol{x}}) = \Psi_m(\boldsymbol{x})(\bar{\boldsymbol{x}}) + \nabla \Psi_m(\boldsymbol{x})(\bar{\boldsymbol{x}})(\boldsymbol{x} - \bar{\boldsymbol{x}}) + \frac{1}{2}(\boldsymbol{x} - \bar{\boldsymbol{x}})' \boldsymbol{C}_m(\bar{\boldsymbol{x}})(\boldsymbol{x} - \bar{\boldsymbol{x}}), \quad \forall \bar{\boldsymbol{x}},$$

where $\boldsymbol{C}_m$ is a $N \times N$ Hermitian nonnegative-definite matrix that ensures (6.5). For simplicity, consider the unstrained estimation case, then given $\bar{\boldsymbol{x}}_1, \cdots, \bar{\boldsymbol{x}}_M$, the minimizer in (6.9) satisfies the following

$$0 = \sum_{m=1}^{M} \left[\nabla \Psi_m(\boldsymbol{x})(\bar{\boldsymbol{x}}_m) + \boldsymbol{C}_m(\bar{\boldsymbol{x}}_m)(\boldsymbol{x}^{new} - \bar{\boldsymbol{x}}_m] \right),$$

which is equivalent to

$$\left[\sum_{m=1}^{M} \boldsymbol{C}_m(\bar{\boldsymbol{x}}_m) \right] \boldsymbol{x}^{new} = \sum_{m=1}^{M} \left[\boldsymbol{C}_m(\bar{\boldsymbol{x}}_m) \bar{\boldsymbol{x}}_m - \nabla \Psi_m(\boldsymbol{x})(\bar{\boldsymbol{x}}_m) \right]. \tag{6.9}$$

The general form of IOT with quadratic surrogates is shown in Table 6.4.

For the case where the cost function is strictly convex, quadratic surrogates ensures the aug-

Table 6.4: IOT algorithm with quadratic surrogates (unconstrained)

Initialize $\{\bar{\boldsymbol{x}}_m : m = 1, \cdots, M\}$
for $n = 1, 2, \cdots,$
  for $m = 1, \cdots, M$

$$\bar{\boldsymbol{x}}_m := \left[\sum_{l=1}^{M} \boldsymbol{C}_l(\bar{\boldsymbol{x}}_l)\right]^{-1} \sum_{l=1}^{M} [\boldsymbol{C}_l(\bar{\boldsymbol{x}}_l)\bar{\boldsymbol{x}}_l - \nabla \Psi_l(\boldsymbol{x})(\bar{\boldsymbol{x}}_l)] \tag{6.10}$$

  end

$$\boldsymbol{x}^{(n)} = \bar{\boldsymbol{x}}_M$$

end

mented function $\bar{\mathsf{F}}$ to be convex with respect to $\boldsymbol{x}$. In this case the minimizer in (6.9) exists, and the algorithm in Table 6.4 will decrease $\bar{\mathsf{F}}$ monotonically. One can establish general conditions that ensure convergence of all of the $\boldsymbol{x}_m$ vectors to $\boldsymbol{x}$ [6].

6.1.2 Accelerated algorithms using double surrogates

We present accelerated OS and IOT algorithms. Our accelerated OS algorithm provides fast initial convergence, but is not a convergent algorithm in general. This motivates us to present an accelerated IOT algorithm that is globally convergent. Both methods can be used sequentially to obtain fast and global convergence.

6.1.2.1 Ordered-subsets with double surrogates

In this section, we present an accelerated OS algorithm. Consider a general PL objective function of the form (2.9) We assume that the data-fit term $\mathsf{L}(\boldsymbol{x})$ has a quadratic surrogate of the form:

$$\mathsf{L}(\boldsymbol{x}) \leq \phi_L(\boldsymbol{x}; \tilde{\boldsymbol{x}}) = \mathsf{L}(\tilde{\boldsymbol{x}}) + \nabla \mathsf{L}(\tilde{\boldsymbol{x}})(\boldsymbol{x} - \tilde{\boldsymbol{x}}) + \frac{1}{2}(\boldsymbol{x} - \tilde{\boldsymbol{x}})' \boldsymbol{D}_L(\tilde{\boldsymbol{x}})(\boldsymbol{x} - \tilde{\boldsymbol{x}}), \quad \forall \tilde{\boldsymbol{x}}, \tag{6.11}$$

with an appropriate diagonal matrix $\boldsymbol{D}$ ˙L. We also assume the regularizer $\boldsymbol{R}(\boldsymbol{x})$ has a quadratic surrogate of the form:

$$\mathsf{R}(\boldsymbol{x}) \leq \phi_R(\boldsymbol{x}; \bar{\boldsymbol{x}}) = \mathsf{R}(\bar{\boldsymbol{x}}) + \nabla \mathsf{R}(\bar{\boldsymbol{x}})(\boldsymbol{x} - \bar{\boldsymbol{x}}) + \frac{1}{2}(\boldsymbol{x} - \bar{\boldsymbol{x}})' \boldsymbol{D}_R(\bar{\boldsymbol{x}})(\boldsymbol{x} - \bar{\boldsymbol{x}}), \quad \forall \bar{\boldsymbol{x}}, \tag{6.12}$$

with an appropriate diagonal matrix $\boldsymbol{D}$ ˙R. Then we define the following "double surrogate" function:

$$\phi(\boldsymbol{x}; \tilde{\boldsymbol{x}}, \bar{\boldsymbol{x}}) \triangleq \phi_L(\boldsymbol{x}; \tilde{\boldsymbol{x}}) + \phi_R(\boldsymbol{x}; \bar{\boldsymbol{x}}). \tag{6.13}$$

By construction, this quadratic surrogate has the following properties:

$$\begin{aligned}\Psi(\boldsymbol{x}) &= \phi(\boldsymbol{x};\boldsymbol{x},\boldsymbol{x}) \\ \Psi(\boldsymbol{x}) &\leq \phi(\boldsymbol{x};\tilde{\boldsymbol{x}},\bar{\boldsymbol{x}}), \quad \forall \tilde{\boldsymbol{x}},\bar{\boldsymbol{x}}.\end{aligned}$$

These properties generalize those of usual optimization transfer methods [60, 65]. For subsequent use, note that from (6.13):

$$\nabla_{\boldsymbol{x}}\,\phi(\boldsymbol{x};\tilde{\boldsymbol{x}},\bar{\boldsymbol{x}}) = \nabla \mathsf{L}(\tilde{\boldsymbol{x}}) + \boldsymbol{D}_L(\tilde{\boldsymbol{x}})(\boldsymbol{x}-\tilde{\boldsymbol{x}}) + \nabla \mathsf{R}(\bar{\boldsymbol{x}}) + \boldsymbol{D}_R(\bar{\boldsymbol{x}})(\boldsymbol{x}-\bar{\boldsymbol{x}}),$$

so the minimizer of the double surrogate function is given as follows:

$$\arg\min_{\boldsymbol{x}} \phi(\boldsymbol{x};\tilde{\boldsymbol{x}},\bar{\boldsymbol{x}}) = \boldsymbol{D}_\phi(\tilde{\boldsymbol{x}},\bar{\boldsymbol{x}})\left(\boldsymbol{D}_L(\tilde{\boldsymbol{x}})\tilde{\boldsymbol{x}} + \boldsymbol{D}_R(\bar{\boldsymbol{x}})\bar{\boldsymbol{x}} - \nabla \mathsf{L}(\tilde{\boldsymbol{x}}) - \nabla \mathsf{R}(\bar{\boldsymbol{x}})\right),$$

where we defined the scaling matrix $\boldsymbol{D}_\phi$ as follows

$$\boldsymbol{D}_\phi(\tilde{\boldsymbol{x}},\bar{\boldsymbol{x}}) = \left[\boldsymbol{D}_L(\tilde{\boldsymbol{x}}) + \boldsymbol{D}_R(\bar{\boldsymbol{x}})\right]^{-1}.$$

A standard optimization transfer algorithm works as follows:

$$\begin{aligned}\boldsymbol{x}^{(n+1)} &= \arg\min_{\boldsymbol{x}} \phi(\boldsymbol{x};\boldsymbol{x}^{(n)},\boldsymbol{x}^{(n)}) \\ &= \boldsymbol{x}^{(n)} - \boldsymbol{D}_\phi(\boldsymbol{x}^{(n)},\boldsymbol{x}^{(n)})\left(\nabla \mathsf{L}(\boldsymbol{x}^{(n)}) + \nabla \mathsf{R}(\boldsymbol{x}^{(n)})\right).\end{aligned}$$

One can show this also decreases Ψ monotonically, i.e., $\Psi(\boldsymbol{x}^{(n+1)}) \leq \Psi(\boldsymbol{x}^{(n)})$. Furthermore, this algorithm converges under suitable conditions [65]. For the case of *separable quadratic surrogates* (*SQS*), usually this type of algorithm converges undesirably slowly [1]. The conventional *ordered-subsets* (*OS*) approach to accelerate convergence is to make the following approximation:

$$\nabla \mathsf{L}(\tilde{\boldsymbol{x}}) \approx \boldsymbol{D}_m \nabla \mathsf{L}_m(\tilde{\boldsymbol{x}}),$$

where L_m is the data-fit term corresponding to the mth subset of the projection views, and $\boldsymbol{D}_m$ is a suitable diagonal matrix, which often simply is $M\boldsymbol{I}$ for M subsets as suggested in (6.3). Using this approximation, we define the following *approximate* surrogate function:

$$\phi_m(\boldsymbol{x};\tilde{\boldsymbol{x}},\bar{\boldsymbol{x}}) \triangleq \mathsf{L}(\tilde{\boldsymbol{x}}) + \boldsymbol{D}_m \nabla \mathsf{L}_m(\tilde{\boldsymbol{x}})(\boldsymbol{x}-\tilde{\boldsymbol{x}}) + \frac{1}{2}(\boldsymbol{x}-\tilde{\boldsymbol{x}})'\boldsymbol{D}_L(\tilde{\boldsymbol{x}})(\boldsymbol{x}-\tilde{\boldsymbol{x}}) + \phi_R(\boldsymbol{x};\bar{\boldsymbol{x}}).$$

For a conventional regularized ordered-subsets method, the minimization step for each subset is given as follows:

$$\begin{aligned} \boldsymbol{x}^{(n,m)} &= \arg\min_{\boldsymbol{x}} \phi_m(\boldsymbol{x}; \boldsymbol{x}^{(n,m-1)}, \boldsymbol{x}^{(n,m-1)}) \\ &= \boldsymbol{x}^{(n,m-1)} - \boldsymbol{D}_\phi(\boldsymbol{x}^{(n,m-1)}, \boldsymbol{x}^{(n,m-1)}) \left(\boldsymbol{D}_m \nabla \text{Ł}\left(\boldsymbol{x}^{(n,m-1)}\right) + \nabla \mathsf{R}\left(\boldsymbol{x}^{(n,m-1)}\right) \right), \end{aligned} \tag{6.14}$$

$$\boldsymbol{x}^{(n,0)} := \boldsymbol{x}^{(n)}, \quad \boldsymbol{x}^{(n+1)} := \boldsymbol{x}^{(n,M)},$$

for $m = 1, \ldots, M$.

This iteration is undesirably slow because it computes the regularization gradient $\nabla \boldsymbol{R}$ for every subset. To reduce this expense, we propose to exploit the generality of the double surrogate (6.13) by using the following novel update:

$$\begin{aligned} \boldsymbol{x}^{(n,m)} &= \arg\min_{\boldsymbol{x}} \phi_m(\boldsymbol{x}; \boldsymbol{x}^{(n,m-1)}, \boldsymbol{x}^{(n)}) \\ &= \boldsymbol{D}_\phi(\boldsymbol{x}^{(n,m-1)}, \boldsymbol{x}^{(n)})[\boldsymbol{D}_L(\boldsymbol{x}^{(n,m-1)})\boldsymbol{x}^{(n,m-1)} + \boldsymbol{D}_R(\boldsymbol{x}^{(n)})\boldsymbol{x}^{(n)} \\ &\qquad - \boldsymbol{D}_m \nabla \text{Ł}\left(\boldsymbol{x}^{(n,m-1)}\right) - \nabla \mathsf{R}(\boldsymbol{x}^{(n)})] \\ &= \boldsymbol{x}^{(n,m-1)} - \boldsymbol{D}_\phi(\boldsymbol{x}^{(n,m-1)}, \boldsymbol{x}^{(n)})[\boldsymbol{D}_m \nabla \text{Ł}\left(\boldsymbol{x}^{(n,m-1)}\right) + \nabla \mathsf{R}(\boldsymbol{x}^{(n)}) \\ &\qquad + \underbrace{\boldsymbol{D}_R(\boldsymbol{x}^{(n)})(\boldsymbol{x}^{(n,m-1)} - \boldsymbol{x}^{(n)})}_{\text{new term}}]. \end{aligned} \tag{6.15}$$

This new iteration (6.15) utilizes the *same* regularizer gradient for all subsets. Compared to (6.14), the updates in (6.15) are quite similar except for an extra term that compensates for not updating the regularizer gradient. Table 6.5 summarizes the proposed algorithm. It requires storing the previous image $\boldsymbol{x}^{\text{last}}$.

Table 6.5: General PL ordered subsets method with double surrogate

Initialize $\boldsymbol{x}^{(0)}$
for $n = 0, 1, \cdots,$

$$\boldsymbol{x}^{(n)} := \boldsymbol{x}^{(n,0)}, \quad \boldsymbol{x}^{\text{last}} := \boldsymbol{x}^{(n)}$$

  for $m = 1, \cdots, M$
    if $\text{mod}(m, U_f) = 0$

$$\boldsymbol{x}^{\text{last}} = \boldsymbol{x}^{(n,m-1))}$$

    end

$$\boldsymbol{x}^{(n,m)} = \arg\min_{\boldsymbol{x}} \phi_m(\boldsymbol{x}; \boldsymbol{x}^{(n,m-1)}, \boldsymbol{x}^{\text{last}})$$

  end

$$\boldsymbol{x}^{(n+1)} := \boldsymbol{x}^{(n,M)}$$

end

In above description, we updated the regularizer gradient only after *all* subsets were updated. Obviously, the regularizer gradient can be updated as frequently as needed and we denote the update frequency as U_f (see algorithm in Table 6.5). Updating the regularizer gradient less often will reduce the computational cost at the expense of the convergence rate in early iterations.

The proposed algorithm was evaluated on a PWLS image reconstruction problem for cone-beam X-ray CT. We have investigated the trade-off between the update frequency and the computational expense per iteration.

6.1.2.2 Incremental optimization transfer method with double surrogates

In this section, we present an accelerated IOT algorithm, which is convergent. For most penalized-likelihood image reconstruction problems, the cost function (3.1) can be rewritten as the following form:

$$\Psi(\boldsymbol{x}) = \sum_{m=1}^{M} \mathsf{L}_m(\boldsymbol{x}) + \mathsf{R}(\boldsymbol{x}),$$

where $\mathsf{L}_m(\boldsymbol{x})$ denotes the negative log-likelihood term associated with the mth data block. Usually evaluating gradients of L_m is much more computationally expensive than that of $\mathsf{R}(\boldsymbol{x})$, but for large M, calculating the gradients of $\mathsf{R}(\boldsymbol{x})$ for each data block update can be computationally exhausting. We again use the double surrogate idea that defines individual quadratic surrogates for the data-fidelity terms and for the regularization, and update the penalty surrogates less frequently.

For the log-likelihood term, we define quadratic surrogates of the following form

$$\phi_m(\boldsymbol{x}; \bar{\boldsymbol{x}}) = \mathsf{L}_m(\bar{\boldsymbol{x}}) + \nabla \mathsf{L}_m(\bar{\boldsymbol{x}})(\boldsymbol{x} - \bar{\boldsymbol{x}}) + \frac{1}{2}(\boldsymbol{x} - \bar{\boldsymbol{x}})' \boldsymbol{C}_m(\bar{\boldsymbol{x}})(\boldsymbol{x} - \bar{\boldsymbol{x}}), \quad \forall \bar{\boldsymbol{x}}, \tag{6.16}$$

where $\boldsymbol{C}_m$ is a Hermitian nonnegative-definite matrix that ensures ϕ_m to majorize L_m. After constructing the following augmented cost function

$$\begin{aligned} \bar{\mathsf{F}}(\boldsymbol{x}; \bar{\boldsymbol{x}}_1, \cdots, \bar{\boldsymbol{x}}_M) &\triangleq \sum_{m=1}^{M} \phi_m(\boldsymbol{x} \| \bar{\boldsymbol{x}}_m) + \mathsf{R}(\boldsymbol{x}), \\ &\equiv \sum_{m=1}^{M} \left(\boldsymbol{x}' \nabla \mathsf{L}_m(\bar{\boldsymbol{x}}_m) + \frac{1}{2}(\boldsymbol{x} - \bar{\boldsymbol{x}}_m)' \boldsymbol{C}_m(\bar{\boldsymbol{x}}_m)(\boldsymbol{x} - \bar{\boldsymbol{x}}_m) \right) + \mathsf{R}(\boldsymbol{x}), \end{aligned} \tag{6.17}$$

we apply alternating minimization instead of minimizing the cost function $\Psi(\boldsymbol{x})$ directly. Minimizing (6.18) over $\boldsymbol{x}$ is essentially a penalized weighted least squares problem. We apply optimization transfer again for $\bar{F}$. Suppose that we have a quadratic surrogate for the regularizer in the form of

(6.12). We define a surrogate function for $\bar{F}$ as follows

$$\Phi(\boldsymbol{x}; \tilde{\boldsymbol{x}}; \bar{\boldsymbol{x}}_1, \cdots, \bar{\boldsymbol{x}}_M) \triangleq \bar{\mathsf{F}}(\boldsymbol{x}; \bar{\boldsymbol{x}}_1, \cdots, \bar{\boldsymbol{x}}_M) - \mathsf{R}(\boldsymbol{x}) + \mathsf{R}(\boldsymbol{x}; \tilde{\boldsymbol{x}}) \,. \tag{6.18}$$

Notice that

$$\arg\min_{\boldsymbol{x}} \bar{F}(\boldsymbol{x}; \cdot) = \arg\min_{\boldsymbol{x}} \underset{\tilde{\boldsymbol{x}}}{min} \Phi(\boldsymbol{x}; \tilde{\boldsymbol{x}}; \cdots).$$

Iteratively minimizing $\Phi(\cdot)$ with respect to $\boldsymbol{x}$ and $\tilde{\boldsymbol{x}}$ will monotonically decrease $\bar{F}$ with respect to $\boldsymbol{x}$. The minimizer of Φ satisfies the following

$$\begin{aligned} 0 &= \nabla_{\boldsymbol{x}} \Phi(\boldsymbol{x}; \tilde{\boldsymbol{x}}; \bar{\boldsymbol{x}}_1, \cdots, \bar{\boldsymbol{x}}_M), \\ &= \sum_{m=1}^{M} \left(\nabla \mathsf{L}_m(\bar{\boldsymbol{x}}_m) + \boldsymbol{C}_m(\bar{\boldsymbol{x}}_m)(\boldsymbol{x} - \bar{\boldsymbol{x}}_m) \right) + \nabla \mathsf{R}(\tilde{\boldsymbol{x}}) + \boldsymbol{D}_R(\tilde{\boldsymbol{x}})(\boldsymbol{x} - \tilde{\boldsymbol{x}}), \\ &= \left[\hat{\boldsymbol{C}} + \boldsymbol{D}_R(\tilde{\boldsymbol{x}}) \right] \boldsymbol{x} - \sum_{m=1}^{M} \left[\boldsymbol{C}_m(\bar{\boldsymbol{x}}_m) \bar{\boldsymbol{x}}_m - \nabla \mathsf{L}_m(\bar{\boldsymbol{x}}_m) \right] + \nabla \mathsf{R}(\tilde{\boldsymbol{x}}) + \boldsymbol{D}_R(\tilde{\boldsymbol{x}}) \tilde{\boldsymbol{x}}, \end{aligned}$$

where $\hat{\boldsymbol{C}} \triangleq \sum_{m=1}^{M} \boldsymbol{C}_m(\bar{\boldsymbol{x}}_m)$. Solving this yields the following inner iteration:

$$\boldsymbol{x}^{\text{new}} := \left[\hat{\boldsymbol{C}} + \boldsymbol{D}_R(\boldsymbol{x}^{\text{old}}) \right]^{-1} \left[\sum_{l=1}^{M} \left[\boldsymbol{C}_l(\bar{\boldsymbol{x}}_l) \bar{\boldsymbol{x}}_l - \nabla \mathsf{L}_l(\bar{\boldsymbol{x}}_l) \right] + \nabla \mathsf{R}\big(\boldsymbol{x}^{\text{old}}\big) + \boldsymbol{D}_R(\boldsymbol{x}^{\text{old}}) \boldsymbol{x}^{\text{old}} \right]. \tag{6.19}$$

Table 6.6 summarizes the proposed algorithm with the update frequency U_f. One extreme case is that we consider the regularization term as a "block" in addition to the data blocks [34]. In this case, the computational cost for the regularizer may become very low, but we may lose "subset gradient balance". The proposed algorithm was also evaluated on a PWLS image reconstruction problem for cone-beam X-ray CT.

6.1.3 Simulation results for sparse view case

The proposed algorithms were investigated on a 3D cone-beam CT image reconstruction problem with limited view angles. The XCAT phantom [112] was used as a static object, and the image was reconstructed to a $512 \times 512 \times 50$ grid with pixel size $\Delta_x = \Delta_y = 0.9766$ mm and $\Delta_z = 0.625$ mm.

We simulated a 3rd-generation axial cone-beam CT system using the separable footprint projector [79]. The simulated system has $N_s = 888$ channels and $N_t = 32$ detector rows spaced by $\Delta_s = 1.0239$ mm and $\Delta_t = 1.09878$ mm per view, and 164 evenly spaced view angles over 360°, which corresponds to an undersampling factor of 6. The source to detector distance was 949 mm, and the source to rotation center distance was 541 mm. We also included a quarter detector offset

Table 6.6: IOT algorithm with quadratic double surrogates for PL problems

Initialize $\{\bar{\boldsymbol{x}}_m : m = 1, \cdots, M\}$ and $\boldsymbol{x}^{(0)} = \boldsymbol{x}^{\text{old}}$

for $n = 1, 2, \cdots,$

$$\bar{\boldsymbol{x}}_0 \coloneqq \boldsymbol{x}^{(n-1)}$$

for $m = 1, \cdots, M$

if $mod(m, U_f)$ = 0

$$\boldsymbol{x}^{\text{old}} = \bar{\boldsymbol{x}}_{m-1} \tag{6.20}$$

end

$$\bar{\boldsymbol{x}}_m \coloneqq \left[\hat{\boldsymbol{C}} + \boldsymbol{D}_R(\boldsymbol{x}^{old})\right]^{-1} \left[\sum_{l=1}^{M} [\boldsymbol{C}_l(\bar{\boldsymbol{x}}_l)\bar{\boldsymbol{x}}_l - \nabla \mathsf{L}_l(\bar{\boldsymbol{x}}_l)] + \nabla \mathsf{R}\big(\boldsymbol{x}^{\text{old}}\big) + \boldsymbol{D}_R(\boldsymbol{x}^{\text{old}})\boldsymbol{x}^{\text{old}}\right] \tag{6.21}$$

end

$$\boldsymbol{x}^{(n+1)} \coloneqq \bar{\boldsymbol{x}}_M$$

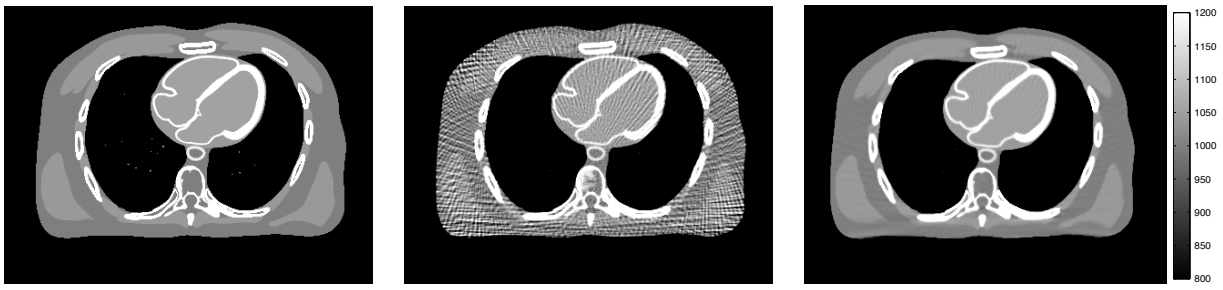
end

in the channel direction to reduce aliasing.

For the edge-preserving regularizer, we used a certainty-based penalty [42] to obtain more uniform resolution and a q-Generalized Gaussian MRF (q-GGMRF) [124] as the penalty function to provide edge-preserving regularizer. The regularization parameter β was selected such that the target PSF has a full-width at half-maximum (FWHM) of approximately 1.4 mm [42]. We generated the noisy sinogram with Poisson noise, and used weighting $w_i = \exp(-[\boldsymbol{A}\boldsymbol{x}]_i)$.

To assess the convergence speed of the proposed method, we computed the root mean squared difference between the image estimate at the nth iteration, $\boldsymbol{x}^{(n)}$, with both the fully converged solution, $\boldsymbol{x}^{\infty}$, and true image, $\boldsymbol{x}^{\text{true}}$. We calculated $\boldsymbol{x}^{\infty}$ using consecutive steps of ordered-subsets with decreasing number of subsets. We used 41, 10, and 1 subsets with 100, 100, and 1000 iterations respectively. This type of "relaxed" OS is guaranteed to converge because the final stage uses just 1 subset for which (6.14) is convergent [1]. However, we also performed 100 iterations of IOT after that to make sure the convergence.

Fig. 6.1 shows the images of true phantom, Filtered Back Projection (FBP) reconstruction [36], and fully converged image ($\boldsymbol{x}^{\infty}$). For the FBP reconstruction, the ramp filter was associated with a Hanning apodization window to attenuate the high frequency noise. Due to limited view angles, conventional FBP reconstruction, which was used as the initial condition for our OS reconstruction, shows severe streaking artifacts compared to the true phantom. The fully converged image has much less artifacts, thus illustrating the benefits of statistical image reconstruction in limited view



(a) (b) (c)

Figure 6.1: Images of (a) XCAT phantom and (b) FBP reconstruction (c) Converged Image ($\boldsymbol{x}^{\infty}$) from cone-beam CT data with 164 projection views.

angle problems.

We divided the projections into 41 subsets, which corresponds to 4 views per subset. This is a rather aggressive selection compared to conventional choices to try to accelerate convergence significantly. The regularizer gradient was updated at different frequencies to see its effect to convergence and computation time.

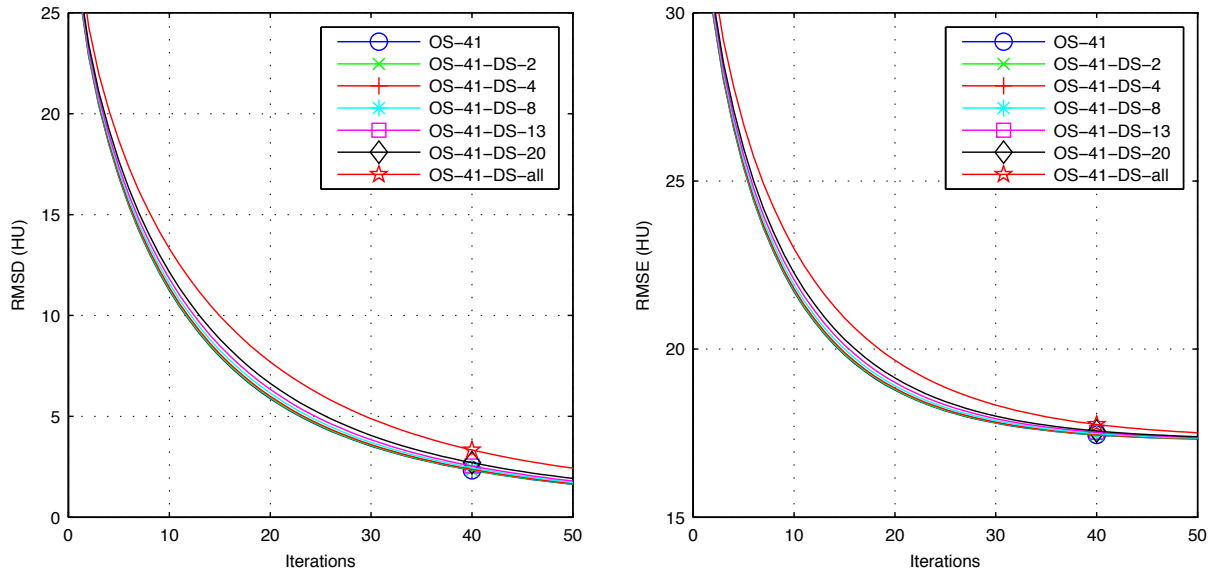
6.1.3.1 OS-DS results

Fig. 6.2 illustrates that our proposed method gives similar root-mean-square (RMS) differences,

$$\sqrt{\frac{1}{N}\sum_{j=1}^{N}(x_j^{(n)} - x_j^{\infty})},$$

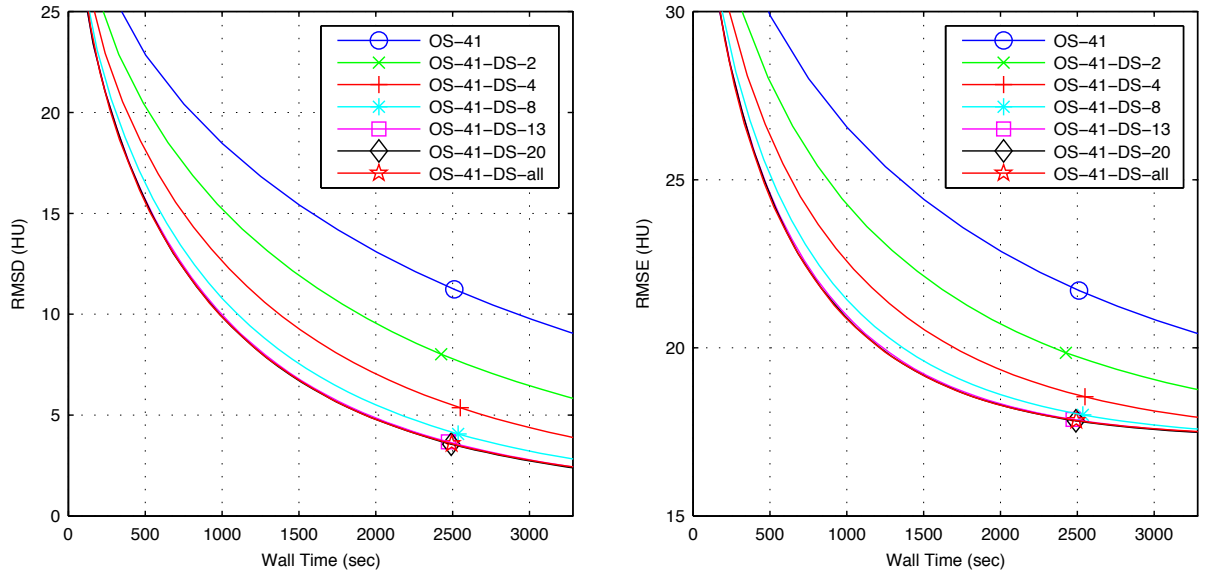
as the conventional OS even when we update the regularizer infrequently per iteration.

On the other hand, Fig. 6.3 illustrates that the computational expense required to obtain the same level of RMS differences was reduced by the proposed method. With $U_f = 13$, which gives the best result in this case, the proposed method converges about 3 times faster than the conventional OS. By observing the reconstructed image at the same time point, Fig. 6.4 clearly shows that the proposed method is converging faster. There exists tradeoff between convergence rates and computational expense, and for our case calculating the regularizer gradient for every 13 subset updates gave the most efficient results. For different problems, the optimal update frequency will differ. However, it is noticeable that regardless of the update frequency, the proposed method is converging faster than the conventional OS. As the problem gets larger and the number of subsets increase, the computational expenses required to calculate the gradient of the regularizer becomes much more dominant. Therefore, we can expect substantial benefits from our proposed method for such problems.



(a) (b)

Figure 6.2: Convergence rates of OS-DS at each iteration for different regularizer update frequency. OS-41-DS-n indicates OS with 41 subsets and U_f = n, and "n = all" means only updating once after all subset updates are done. (a) $\boldsymbol{x}^{(n)}$ with respect to $\boldsymbol{x}^{\infty}$ and (b) $\boldsymbol{x}^{(n)}$ with respect to $\boldsymbol{x}^{\text{true}}$



(a) (b)

Figure 6.3: Convergence rates of OS-DS versus time for different regularizer update frequency. (a) $\boldsymbol{x}^{(n)}$ with respect to $\boldsymbol{x}^{\infty}$ and (b) $\boldsymbol{x}^{(n)}$ with respect to $\boldsymbol{x}^{\text{true}}$

To see the effects of the new term in (6.16), we compare the reconstructed images obtained with and without the term (see Fig. 6.5). Without the new update term, less frequent update of regularizer gradient led to very poor convergence characteristics. The resulting image is severely corrupted in the unsampled region of the image (see Fig. 6.6).

6.1.3.2 IOT-DS results

We used 1 iteration of OS-SQS algorithm to initialize IOT [6].

Comparing the results for OS-DS (Fig. 6.2 and Fig. 6.3) and that of IOT-DS (Fig. 6.7 and Fig. 6.8), we can observe similar tendencies. Less frequent regularizer gradient updates have very little effects on the convergence in iterations Fig. 6.7. On the other hand, Fig. 6.8 shows that the same level of RMS differences can be obtained with much less computation by using the proposed IOT with double surrogates. In this case, regarding the regularization as one of the block performed most efficiently, and we obtained approximately 2.5 times speed up. For different problems, the optimal update frequency may differ, but same as the OS-DS case, the IOT-DS method converges faster regardless of the update frequency.

Since 1 iterations of OS-SQS was not enough to reach the limit cycle, we can notice the undesirable jump in the convergence curve [6].

6.1.4 Physical phantom results

We reconstructed $512 \times 512 \times 47$ images of the GEPP from the measurements with 888 channels, 64 rows, and 3071 views. It was a helical scan with pitch 0.5. We used OS-DS algorithm with 328 subsets and the scaling factor suggested in [74] to reconstruct images.

In Table 6.5, we suggested to update regularizer gradients for every U_f inner iterations. We also investigated different update patterns. We use the number of updates N_u instead to indicate the frequency of regularizer gradient updates. Here are possible three patterns:

- "Regular" - This corresponds to the current update pattern.
- "Consecutive" - Do consecutive N_u updates at the end of the inner iteration.
- "Random-1" : Randomly do $N_u - 1$ updates during the inner iteration, and always update at the last inner iteration.

Fig. 6.9 compares the effect of these patterns to convergence speed. Number of updates N_u was fixed to 10 for the experiment. Consecutively updating the regularizer gradient at the end of inner iterations turns out to be a bad strategy. On the other hand, the other two update patterns showed results. It seems that the reason is because we used uniform distribution to determine when to update. Using different probability distribution may change the results.

Fig. 6.10 illustrates how the number of updates N_u affects the convergence speed in both iterations and wall time. Similar to results obtained with the sparse view simulation, the proposed method reduced the computational cost to reach the same RMSD level. We obtained about 2 to 3 times speed up for this case. However, if we reduce the number of updates too much, we severely lose the convergence rates, which is undesirable.

6.1.5 Simulation results for low-dose large cone-angle case

We simulated a 256-slice axial cone-beam CT scanner and a low-dose case with 50mA tube current. We obtained the measurements with 888 channels, 256 rows, and 622 views, and XCAT phantom was used as a static object. The image was reconstructed to a $718 \times 718 \times 440$ grid with the same voxel size as the previous simulation case. We applied double surrogate method to OS with Nesterov's fast gradient method [75] (OS-NES-DS), and used it to reconstruct images. We modified $\boldsymbol{x}$ update equation in Fig. 2 of [75] by applying (6.16), so that the gradient of the regularizer is less frequently updated. We used 12 subsets for the algorithm (OS-NES-12-DS).

Fig. 6.11 shows that the proposed method, OS-NES-12-DS, further accelerates the OS + momentum method in time. Using small number of updates leads to decreased convergence rates in iteration, but requires much less computation per iteration. The main reason for the speed up is the large size of the reconstruction volume. Even for small number of subsets, the regularizer gradient updates can be computationally expensive.

6.1.6 Conclusions

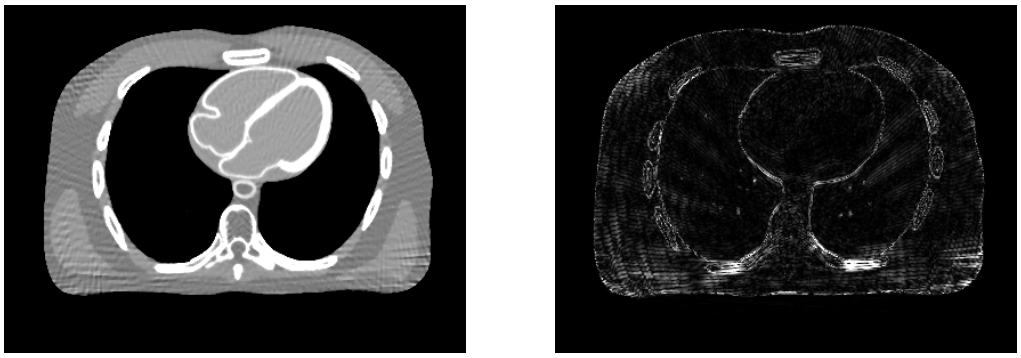
We presented simple modifications of the OS type methods that allow the regularizer gradient to be less frequently updated. The methods provide very significant acceleration when applied to large problems with many subsets or huge reconstruction volume, such as large cone-angle cone-beam CT or helical CT image reconstruction, while still providing good reconstructed images.

Simulations and phantom experiments demonstrated that a good reconstruction with compatible quality was achieved within much less computation time. However, the reduction in the computational expense may depend on the size of the problem and the complexity of updating the regularizer compared to the complexity of forward and back-projection. When implemented on graphics processing unit (GPU), the trade-offs may differ.

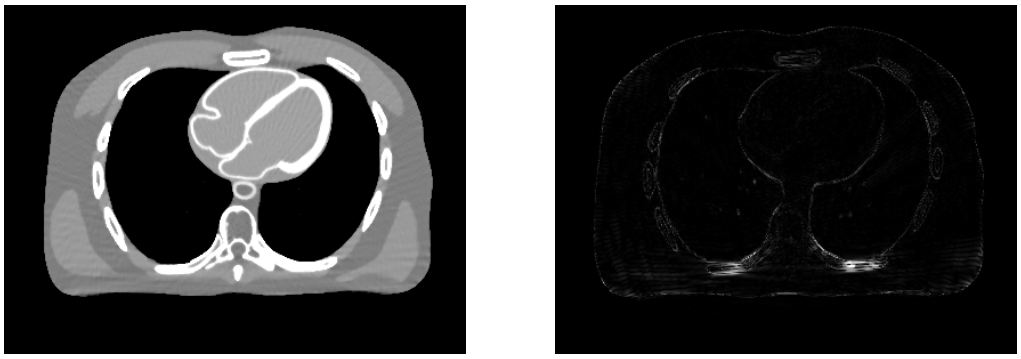
For ordered-subsets method, we focussed on the initial convergence characteristics of the method. Since the conventional OS algorithm is inherently not globally convergent, neither is the proposed method. Thus we also applied the double surrogate idea to incremental optimization transfer method, which is a globally convergent block iterative algorithm. Switching from OS-DS to IOT-DS will give faster initial convergence and also convergence to the solution. Finally, we

applied the idea to one of the recent variations of OS algorithm, OS + momentum, and showed its effectiveness in reducing computation time.

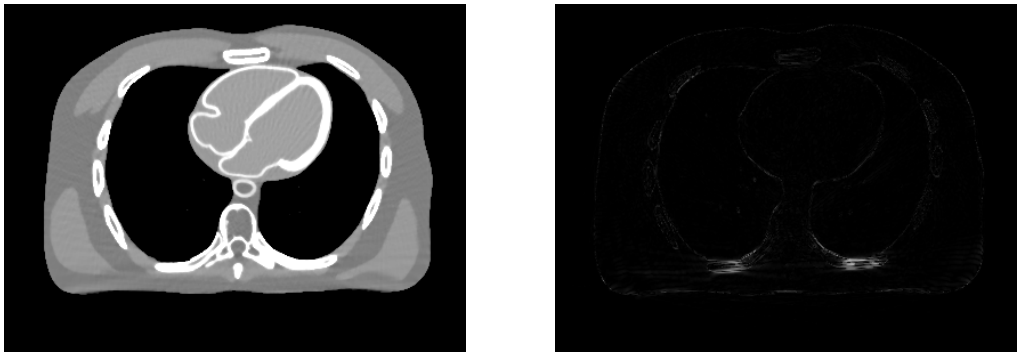
One issue with the double surrogate method, as mentioned before, is the selection of the update frequency. Using moderate update frequency can provide reasonable speed up, but we want to maximize it. Further research will address the systematic way to determine the optimal update frequency and possible improvements of the methods.



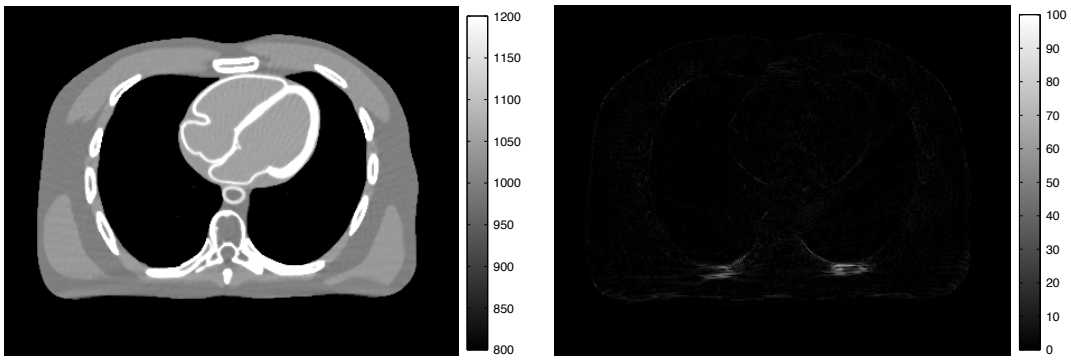
(a) OS-41 (conventional OS)



(b) OS-41-DS-4



(c) OS-41-DS-13



(d) OS-41-DS-all

Figure 6.4: Comparing the convergence speed of OS-DS with different update frequencies. Left column: Images at the same time point (4000 sec after initialization). Right column: Absolute difference images with respect to $\boldsymbol{x}^{(\infty)}$.

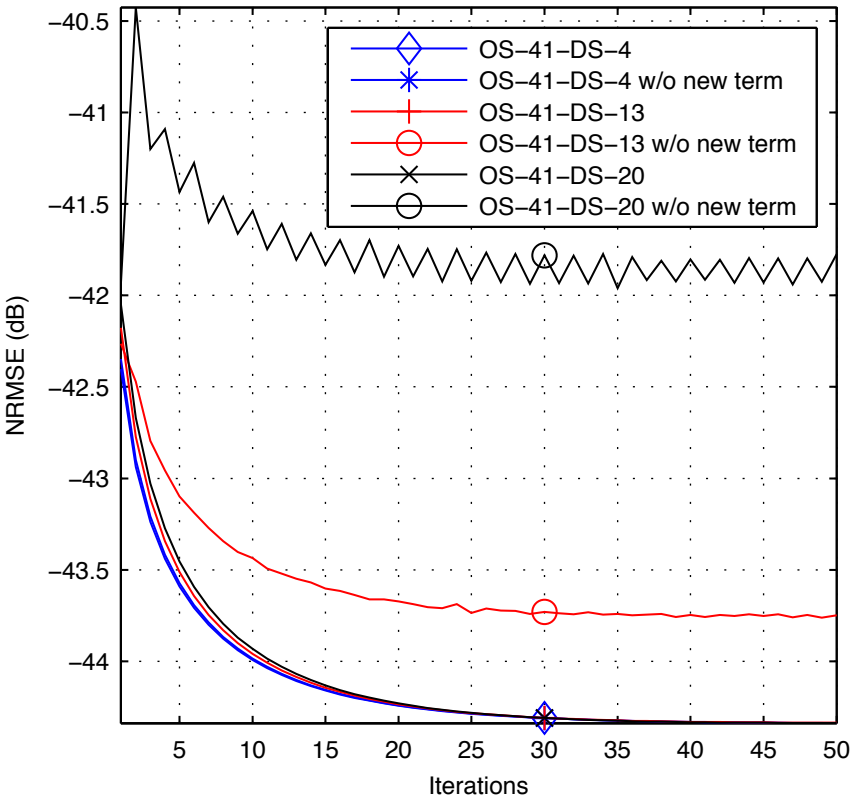


Figure 6.5: Effects of the new update term in the convergence rate

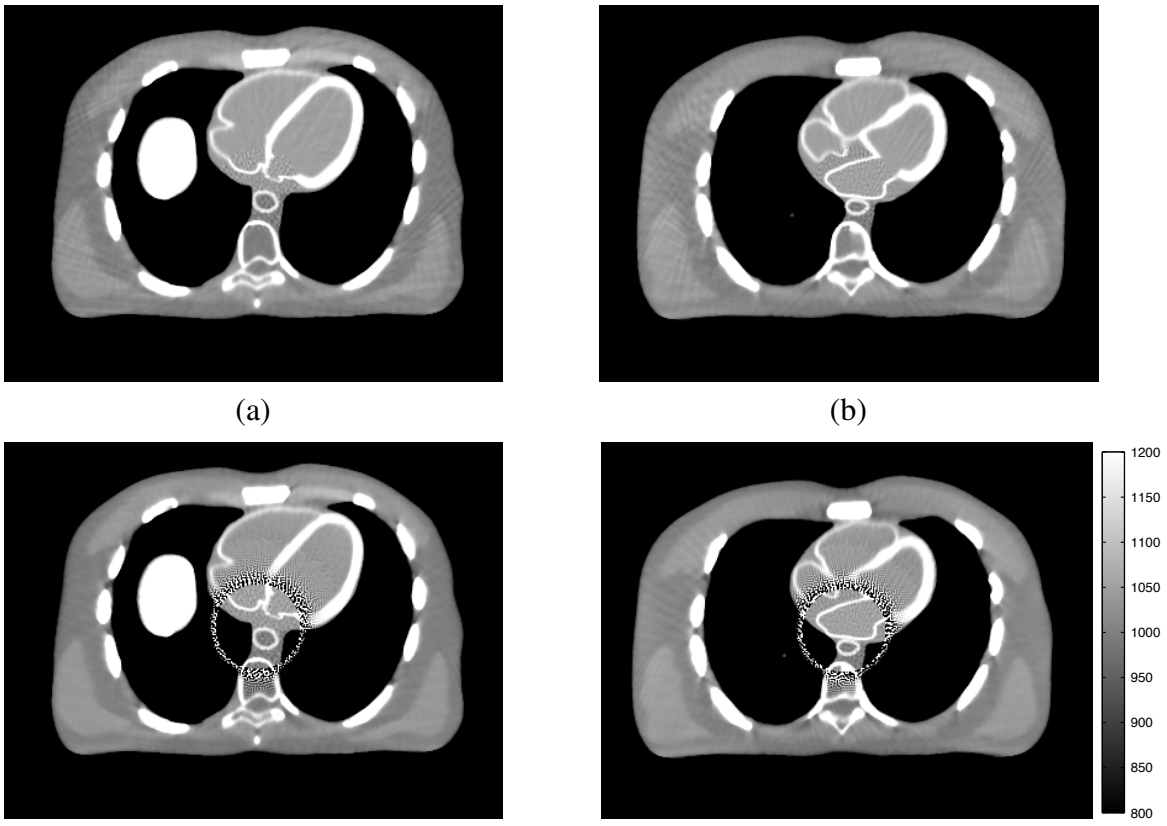
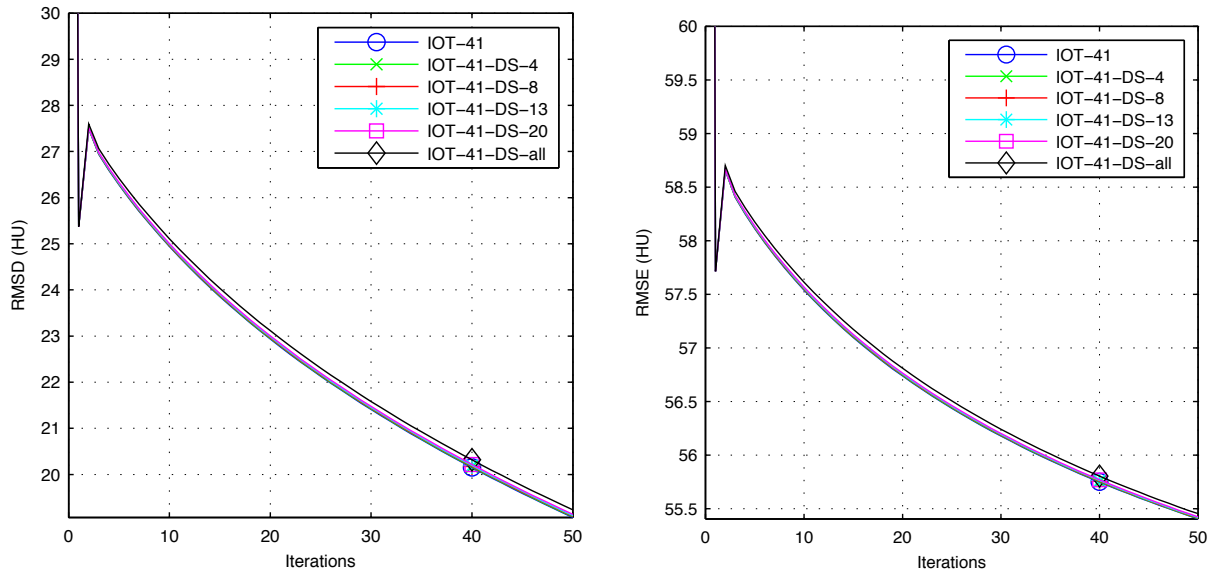
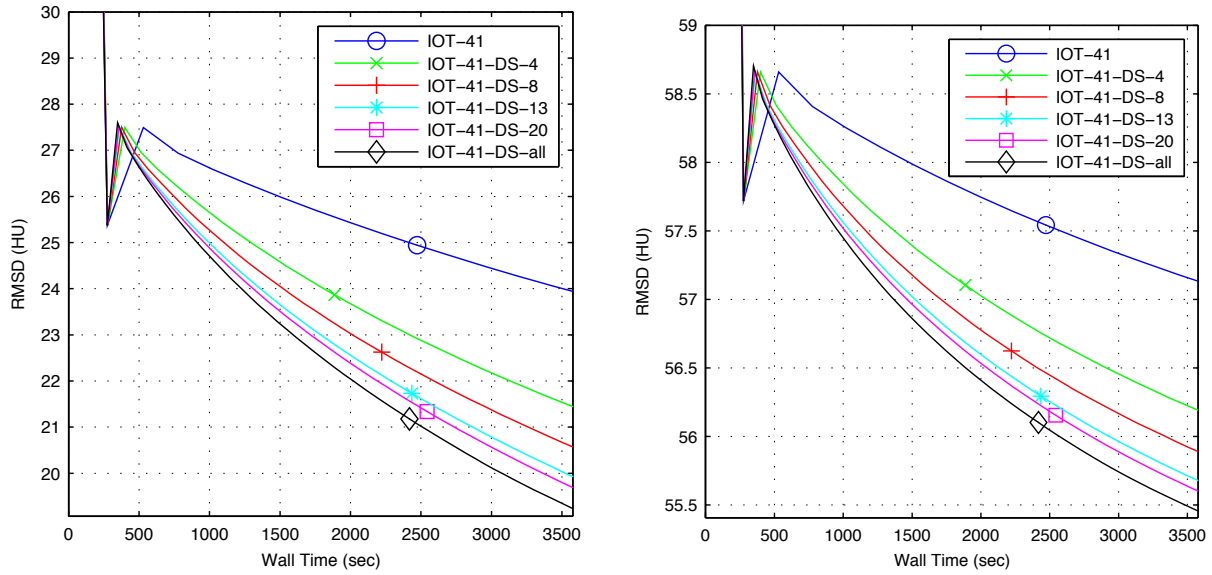


Figure 6.6: Images of OS-41-DS-all with and without the new update term (a) With the new term, two slice below bottom end slice (b) With the new term, two slice above top end slice (c) Without the new term, two slice below bottom end slice (d) Without the new term, two slice above top end slice



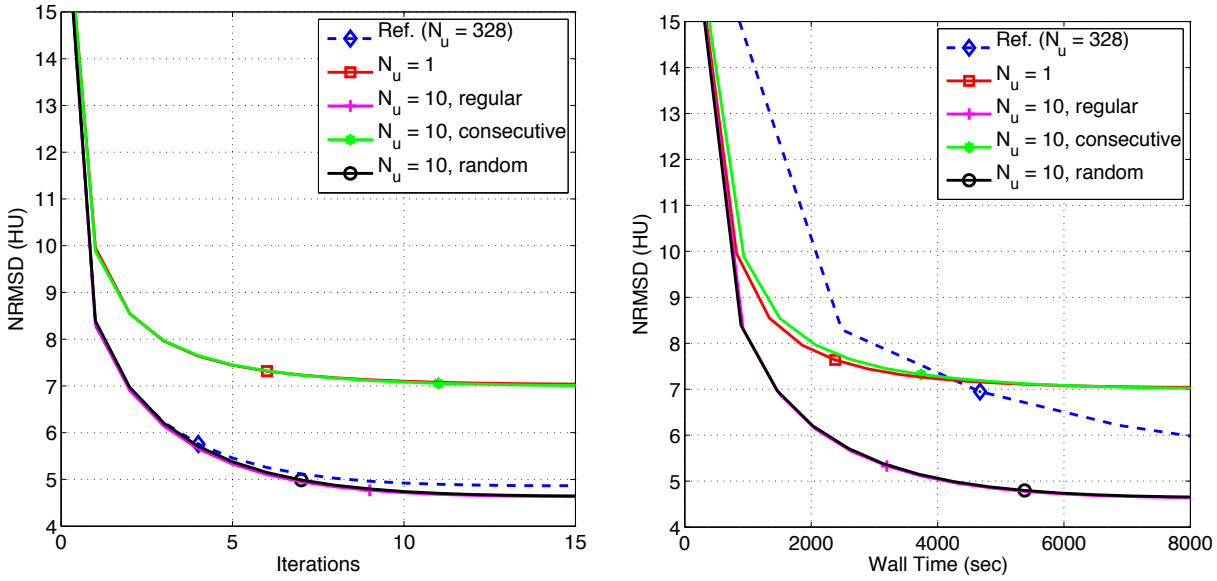
(a) (b)

Figure 6.7: Convergence rates of IOT-DS at each iteration for different regularizer update frequency. IOS-41-DS-n indicates IOT with 41 subsets and U_f = n, and "n = all" means only updating once after all subset updates are done. (a) $\boldsymbol{x}^{(n)}$ with respect to $\boldsymbol{x}^{\infty}$ and (b) $\boldsymbol{x}^{(n)}$ with respect to $\boldsymbol{x}^{\text{true}}$



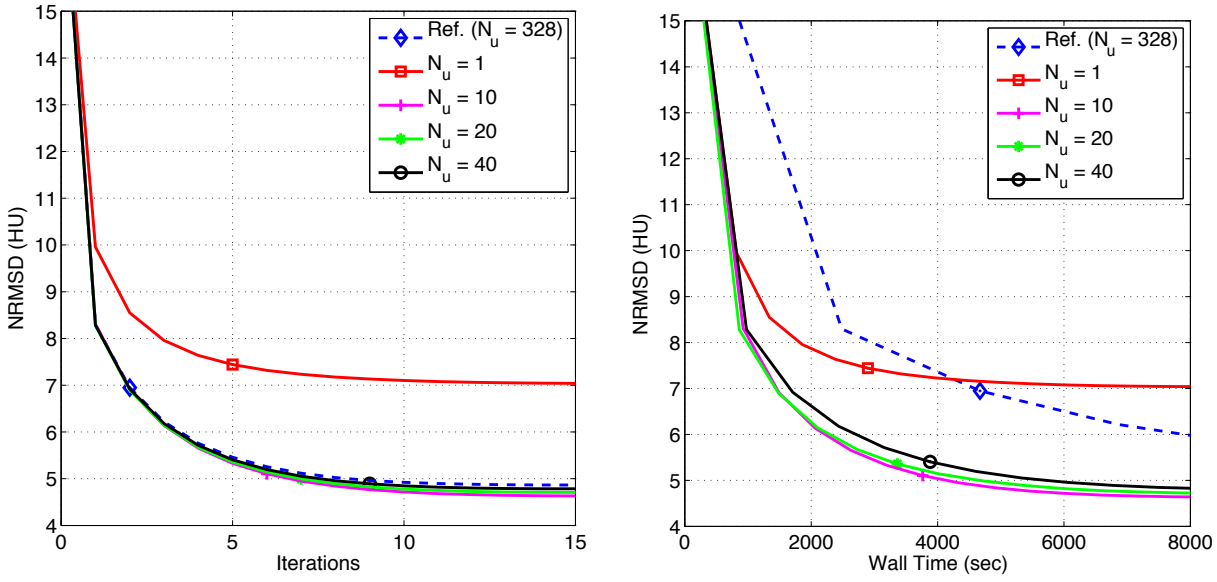
(a) (b)

Figure 6.8: Convergence rates of IOT-DS versus time for different regularizer update frequency. (a) $\boldsymbol{x}^{(n)}$ with respect to $\boldsymbol{x}^{\infty}$ and (b) $\boldsymbol{x}^{(n)}$ with respect to $\boldsymbol{x}^{\text{true}}$



(a) (b)

Figure 6.9: Convergence rates of OS-DS for different update patterns (a) versus iterations (b) versus time (sec).



(a) (b)

Figure 6.10: Convergence rates of OS-DS for different number of updates (N_u) (a) versus iterations (b) versus time (sec).

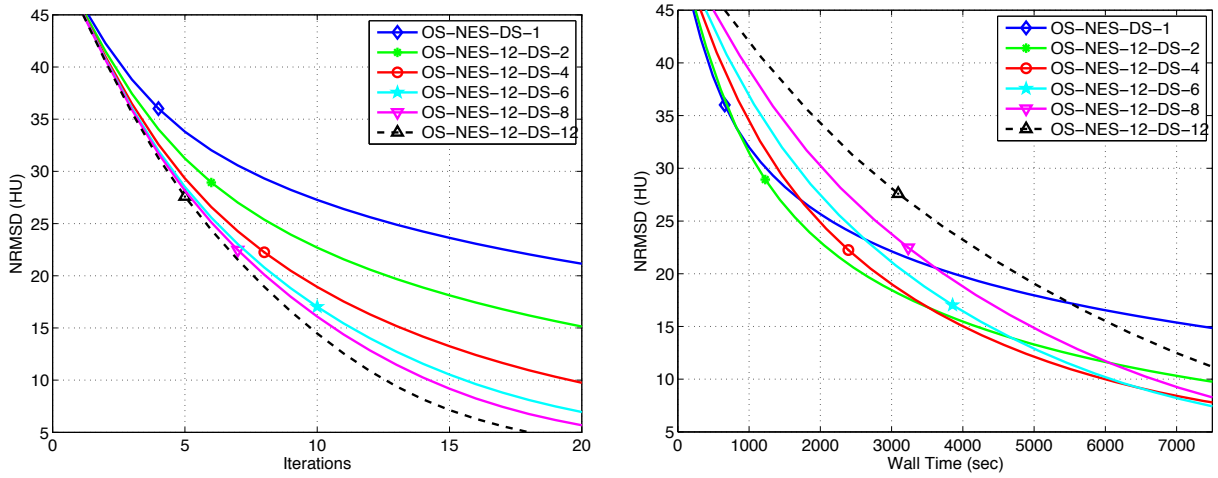


Figure 6.11: Convergence rates of OS-NES-12-DS-N_u for different number of updates (N_u) (a) versus iterations (b) versus time (sec).

6.2 Motion-compensated image reconstruction with alternating minimization

Cardiac computed tomography (CT) is important for diagnosing heart disease. Motion artifacts are a significant issue for cardiac CT image reconstruction. Motion-compensated image reconstruction (MCIR) has the potential to overcome the drawbacks of conventional gated reconstruction methods by exploiting all the measurement data and using motion information. In general, MCIR methods consist of two main steps: estimating the motion and reconstructing the image using the estimated motion. The quality of the reconstructed image is significantly affected by the accuracy of the estimated motion, and thus many researchers have focussed on improving motion estimates. However, the image reconstruction part is also very important for practical use of MCIR methods. Since the system model in MCIR methods has both the forward-projector and the warp matrices, it becomes computationally very expensive to use iterative algorithms for MCIR. Unlike conventional CT image reconstruction problems, designing a proper preconditioner for MCIR is not trivial due to the complexity of the system model. Ordered-subset (OS) type of algorithms are not efficient for MCIR, especially when when warping is computationally expensive.

In this chapter, we propose a novel approach to solve the image reconstruction part of the MCIR method more efficiently. We use a variable-splitting technique to dissociate the original problem into a number of simpler problems. The proposed method is amenable to preconditioning, parallelization, and application of block iterative algorithms to sub-problems. We demonstrated through a phantom simulation that with simple diagonal or circulant preconditioners, the proposed method shows good convergence rate compared to conjugate gradient (CG) method.

6.2.1 Motion-compensated image reconstruction for CT

6.2.1.1 Problem formulation

We use the mathematical models defined in Chapter II. Consider a penalized-likelihood least squares (PWLS) formulation of motion-compensated CT image reconstruction:

$$\hat{\boldsymbol{x}} = \underset{\boldsymbol{x}}{arg\,min} \left\{ \Psi(\boldsymbol{x}) \triangleq \mathsf{L}(\boldsymbol{x}) + \mathsf{R}(\boldsymbol{C}\boldsymbol{x}) \right\}, \tag{6.22}$$

$$\mathsf{L}(\boldsymbol{x}) = \frac{1}{2} \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{T}\boldsymbol{x}\|_{\boldsymbol{W}}^2, \quad \mathsf{R}(\boldsymbol{C}\boldsymbol{x}) = \beta \sum_{k=1}^{K} \kappa_k \, \psi_k([\boldsymbol{C}\boldsymbol{x}]_k),$$

$$\boldsymbol{A} = \mathrm{diag}\{\boldsymbol{A}_1, \cdots, \boldsymbol{A}_{N_f}\}, \quad \boldsymbol{T} = [\boldsymbol{T}_1' \ldots \boldsymbol{T}_{N_f}']',$$

where $\boldsymbol{A}_m$ is the system matrix corresponding to mth frame, β_k is the regularization parameter, ψ_k is the potential function, $\boldsymbol{C}$ is a matrix that performs finite differences between neighboring

voxels, K is the number of neighbors, and $\boldsymbol{T}$ is the warp matrix based on motion parameters that are assumed to be known. The minimization problem (6.22) is challenging due to the warp matrix $\boldsymbol{T}$ in the system model.

6.2.2 Proposed method

We apply a variable splitting approach to the problem. The basic idea of variable splitting method is to introduce auxiliary constraint variables so that coupled parts in the cost function can be separated [3, 46]. The original problem is transformed into an equivalent constrained optimization problem, and then alternating minimization methods are applied to efficiently solve the problem. Previous works have focussed on splitting the regularization term and also the statistical weighting [3, 103, 105]. In this work, in addition to those splittings, we focus on splitting the warp matrix from the forward-projector in the system matrix.

6.2.2.1 Equivalent constrained optimization problem

We introduce auxiliary constraint variables $\boldsymbol{u}$, $\boldsymbol{v}$, $\boldsymbol{z}$, and $\boldsymbol{s}$, and write (6.22) as the following equivalent constrained problem:

$$\begin{aligned} &\underset{\boldsymbol{x},\boldsymbol{u},\boldsymbol{v},\boldsymbol{z},\boldsymbol{s}}{arg\,min}\, \Psi(\boldsymbol{x},\boldsymbol{u},\boldsymbol{v},\boldsymbol{z},\boldsymbol{s}) = \frac{1}{2}\|\boldsymbol{y}-\boldsymbol{v}\|_{\boldsymbol{W}}^2 + \mathsf{R}(\boldsymbol{z}), \\ &s.t. \quad \boldsymbol{u}=\boldsymbol{T}\boldsymbol{x}, \quad \boldsymbol{v}=\boldsymbol{A}\boldsymbol{u}, \quad \boldsymbol{z}=\boldsymbol{C}\boldsymbol{s}, \quad \boldsymbol{s}=\boldsymbol{x}, \end{aligned} \tag{6.23}$$

where $\boldsymbol{u} \in \mathbb{R}^{NN_f}$ separates the system matrix from the warp matrix, $\boldsymbol{v} \in \mathbb{R}^M$ separates the effect of the weighting matrix, $\boldsymbol{W}$, on $\boldsymbol{A}\boldsymbol{x}$, $\boldsymbol{z} \in \mathbb{R}^{NK}$ and $\boldsymbol{s} \in \mathbb{R}^N$ detach the warp matrix from the regularizer.

6.2.2.2 Method of multipliers

We used the framework of method of multipliers [9, 89] to solve (6.23), and constructed an augmented Lagrangian function as follows [9, 89, 101, 102]:

$$\begin{aligned} L(\boldsymbol{x},\boldsymbol{u},\boldsymbol{v},\boldsymbol{z},\boldsymbol{s}) \;&\triangleq\; \frac{1}{2}\|\boldsymbol{y}-\boldsymbol{v}\|_{\boldsymbol{W}}^2 + \mathsf{R}(\boldsymbol{z}) \\ &+\; \gamma_{\boldsymbol{u}}^T(\boldsymbol{u}-\boldsymbol{T}\boldsymbol{x}) + \frac{\mu_{\boldsymbol{u}}}{2}\|\boldsymbol{u}-\boldsymbol{T}\boldsymbol{x}\|_{\Lambda_{\boldsymbol{u}}}^2 + \gamma_{\boldsymbol{v}}^T(\boldsymbol{v}-\boldsymbol{A}\boldsymbol{u}) + \frac{\mu_{\boldsymbol{v}}}{2}\|\boldsymbol{v}-\boldsymbol{A}\boldsymbol{u}\|_{\Lambda_{\boldsymbol{v}}}^2 \\ &+\; \gamma_{\boldsymbol{z}}^T(\boldsymbol{z}-\boldsymbol{C}\boldsymbol{s}) + \frac{\mu_{\boldsymbol{z}}}{2}\|\boldsymbol{z}-\boldsymbol{C}\boldsymbol{s}\|_{\Lambda_{\boldsymbol{z}}}^2 + \gamma_{\boldsymbol{s}}^T(\boldsymbol{s}-\boldsymbol{x}) + \frac{\mu_{\boldsymbol{s}}}{2}\|\boldsymbol{s}-\boldsymbol{x}\|_{\Lambda_{\boldsymbol{s}}}^2 \end{aligned}$$

where $\gamma_{\boldsymbol{u}} \in \mathbb{R}^{NN_f}$, $\gamma_{\boldsymbol{v}} \in \mathbb{R}^M$, $\gamma_{\boldsymbol{z}} \in \mathbb{R}^{NK}$, and $\gamma_{\boldsymbol{s}} \in \mathbb{R}^N$ are the Lagrange multipliers, $\mu_{\boldsymbol{u}} > 0$, $\mu_{\boldsymbol{v}} > 0$, $\mu_{\boldsymbol{z}} > 0$, and $\mu_{\boldsymbol{s}} > 0$ are the AL penalty parameters, and $\Lambda_{\boldsymbol{u}} \succ 0$, $\Lambda_{\boldsymbol{v}} \succ 0$, $\Lambda_{\boldsymbol{z}} \succ 0$, and

$\Lambda_{\boldsymbol{s}} \succ 0$ are symmetric weighting matrixes. The multiplier term can be absorbed into the quadratic penalty term, and by ignoring the irrelevant constants we obtain:

$$\begin{aligned} L(\boldsymbol{x},\boldsymbol{u},\boldsymbol{v},\boldsymbol{z},\boldsymbol{s}) = \frac{1}{2}\|\boldsymbol{y}-\boldsymbol{v}\|^2_{\boldsymbol{W}} &+ \mathsf{R}(\boldsymbol{z}) \\ &+ \frac{\mu_{\boldsymbol{u}}}{2}\|\boldsymbol{u}-\boldsymbol{T}\boldsymbol{x}-\eta_{\boldsymbol{u}}\|^2_{\Lambda_{\boldsymbol{u}}} + \frac{\mu_{\boldsymbol{v}}}{2}\|\boldsymbol{v}-\boldsymbol{A}\boldsymbol{u}-\eta_{\boldsymbol{v}}\|^2_{\Lambda_{\boldsymbol{v}}} \\ &+ \frac{\mu_{\boldsymbol{z}}}{2}\|\boldsymbol{z}-\boldsymbol{C}\boldsymbol{s}-\eta_{\boldsymbol{z}}\|^2_{\Lambda_{\boldsymbol{z}}} + \frac{\mu_{\boldsymbol{s}}}{2}\|\boldsymbol{s}-\boldsymbol{x}-\eta_{\boldsymbol{s}}\|^2_{\Lambda_{\boldsymbol{s}}} \end{aligned} \tag{6.24}$$

where η's are Lagrange-multiplier-like vectors defined as $\eta_{\boldsymbol{u}} = -\frac{1}{\mu_{\boldsymbol{u}}}\Lambda_{\boldsymbol{u}}^{-1}\gamma_{\boldsymbol{u}}$, $\eta_{\boldsymbol{v}} = -\frac{1}{\mu_{\boldsymbol{v}}}\Lambda_{\boldsymbol{v}}^{-1}\gamma_{\boldsymbol{v}}$, $\eta_{\boldsymbol{z}} = -\frac{1}{\mu_{\boldsymbol{z}}}\Lambda_{\boldsymbol{z}}^{-1}\gamma_{\boldsymbol{z}}$, and $\eta_{\boldsymbol{s}} = -\frac{1}{\mu_{\boldsymbol{s}}}\Lambda_{\boldsymbol{s}}^{-1}\gamma_{\boldsymbol{s}}$. We use $\Lambda_{\boldsymbol{u}} = \boldsymbol{I}_{NN_f}$, $\Lambda_{\boldsymbol{v}} = \boldsymbol{I}_M$, $\Lambda_{\boldsymbol{z}} = \boldsymbol{I}_{NK}$, and $\Lambda_{\boldsymbol{s}} = \boldsymbol{I}_N$, which is a standard approach [4, 43].

To solve the problem (6.23), we alternate between a joint-minimization step and an update step, which is the classical AL scheme [104]:

$$(\boldsymbol{x}^{(j+1)},\boldsymbol{u}^{(j+1)},\boldsymbol{v}^{(j+1)},\boldsymbol{z}^{(j+1)},\boldsymbol{s}^{(j+1)}) = \underset{\boldsymbol{x},\boldsymbol{u},\boldsymbol{v},\boldsymbol{z}}{arg\,min}\, L(\boldsymbol{x},\boldsymbol{u},\boldsymbol{v},\boldsymbol{z},\boldsymbol{s},\eta_{\boldsymbol{u}}^{(j)},\eta_{\boldsymbol{v}}^{(j)},\eta_{\boldsymbol{z}}^{(j)},\eta_{\boldsymbol{s}}^{(j)}) \tag{6.25}$$

$$\eta_{\boldsymbol{u}}^{(j+1)} = \eta_{\boldsymbol{u}}^{(j)} - (\boldsymbol{u}^{(j+1)} - \boldsymbol{T}\boldsymbol{x}^{(j+1)}), \tag{6.26}$$

$$\eta_{\boldsymbol{v}}^{(j+1)} = \eta_{\boldsymbol{v}}^{(j)} - (\boldsymbol{v}^{(j+1)} - \boldsymbol{A}\boldsymbol{u}^{(j+1)}), \tag{6.27}$$

$$\eta_{\boldsymbol{z}}^{(j+1)} = \eta_{\boldsymbol{z}}^{(j)} - (\boldsymbol{z}^{(j+1)} - \boldsymbol{C}\boldsymbol{s}^{(j+1)}), \tag{6.28}$$

$$\eta_{\boldsymbol{s}}^{(j+1)} = \eta_{\boldsymbol{s}}^{(j)} - (\boldsymbol{s}^{(j+1)} - \boldsymbol{x}^{(j+1)}) \tag{6.29}$$

6.2.2.3 Alternating direction minimization

Solving the joint minimization step (6.25) is computationally expensive, so we apply alternating minimization [105] as follows:

$$\boldsymbol{x}^{(j+1)} = \underset{\boldsymbol{x}}{arg\,min}\, L(\boldsymbol{x},\boldsymbol{u}^{(j)},\boldsymbol{v}^{(j)},\boldsymbol{z}^{(j)},\boldsymbol{s}^{(j)},\eta_{\boldsymbol{u}}^{(j)},\eta_{\boldsymbol{v}}^{(j)},\eta_{\boldsymbol{z}}^{(j)},\eta_{\boldsymbol{s}}^{(j)}), \tag{6.30}$$

$$\boldsymbol{u}^{(j+1)} = \underset{\boldsymbol{u}}{arg\,min}\, L(\boldsymbol{x}^{(j+1)},\boldsymbol{u},\boldsymbol{v}^{(j)},\boldsymbol{z}^{(j)},\boldsymbol{s}^{(j)},\eta_{\boldsymbol{u}}^{(j)},\eta_{\boldsymbol{v}}^{(j)},\eta_{\boldsymbol{z}}^{(j)},\eta_{\boldsymbol{s}}^{(j)}), \tag{6.31}$$

$$\boldsymbol{v}^{(j+1)} = \underset{\boldsymbol{v}}{arg\,min}\, L(\boldsymbol{x}^{(j+1)},\boldsymbol{u}^{(j+1)},\boldsymbol{v},\boldsymbol{z}^{(j)},\boldsymbol{s}^{(j)},\eta_{\boldsymbol{u}}^{(j)},\eta_{\boldsymbol{v}}^{(j)},\eta_{\boldsymbol{z}}^{(j)},\eta_{\boldsymbol{s}}^{(j)}), \tag{6.32}$$

$$\boldsymbol{z}^{(j+1)} = \underset{\boldsymbol{z}}{arg\,min}\, L(\boldsymbol{x}^{(j+1)},\boldsymbol{u}^{(j+1)},\boldsymbol{v}^{(j+1)},\boldsymbol{z},\boldsymbol{s}^{(j)},\eta_{\boldsymbol{u}}^{(j)},\eta_{\boldsymbol{v}}^{(j)},\eta_{\boldsymbol{z}}^{(j)},\eta_{\boldsymbol{s}}^{(j)}), \tag{6.33}$$

$$\boldsymbol{s}^{(j+1)} = \underset{\boldsymbol{s}}{arg\,min}\, L(\boldsymbol{x}^{(j+1)},\boldsymbol{u}^{(j+1)},\boldsymbol{v}^{(j+1)},\boldsymbol{z}^{(j+1)},\boldsymbol{s},\eta_{\boldsymbol{u}}^{(j)},\eta_{\boldsymbol{v}}^{(j)},\eta_{\boldsymbol{z}}^{(j)},\eta_{\boldsymbol{s}}^{(j)}). \tag{6.34}$$

At the jth iteration, we update each vector in turn as follows:

$$\begin{aligned}
\boldsymbol{x}^{(j+1)} &= \underset{\boldsymbol{x}}{arg\,min}\, \frac{\mu_{\boldsymbol{u}}}{2}\left\|\boldsymbol{u}^{(j)} - \boldsymbol{T}\boldsymbol{x} - \eta_{\boldsymbol{u}}^{(j)}\right\|_{\Lambda_{\boldsymbol{u}}}^2 + \frac{\mu_{\boldsymbol{s}}}{2}\left\|\boldsymbol{s}^{(j)} - \boldsymbol{x} - \eta_{\boldsymbol{s}}^{(j)}\right\|_{\Lambda_{\boldsymbol{s}}}^2, && (6.35)\\
\boldsymbol{u}^{(j+1)} &= \underset{\boldsymbol{u}}{arg\,min}\, \frac{\mu_{\boldsymbol{u}}}{2}\left\|\boldsymbol{u} - \boldsymbol{T}\boldsymbol{x}^{(j+1)} - \eta_{\boldsymbol{u}}^{(j)}\right\|_{\Lambda_{\boldsymbol{u}}}^2 + \frac{\mu_{\boldsymbol{v}}}{2}\left\|\boldsymbol{v}^{(j)} - \boldsymbol{A}\boldsymbol{u} - \eta_{\boldsymbol{v}}^{(j)}\right\|_{\Lambda_{\boldsymbol{v}}}^2, && (6.36)\\
\boldsymbol{v}^{(j+1)} &= \underset{\boldsymbol{v}}{arg\,min}\, \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{v}\|_{\boldsymbol{W}}^2 + \frac{\mu_{\boldsymbol{v}}}{2}\left\|\boldsymbol{v} - \boldsymbol{A}\boldsymbol{u}^{(j+1)} - \eta_{\boldsymbol{v}}^{(j)}\right\|_{\Lambda_{\boldsymbol{v}}}^2, && (6.37)\\
\boldsymbol{s}^{(j+1)} &= \underset{\boldsymbol{s}}{arg\,min}\, \frac{\mu_{\boldsymbol{z}}}{2}\left\|\boldsymbol{z}^{(j)} - \boldsymbol{C}\boldsymbol{s} - \eta_{\boldsymbol{z}}^{(j)}\right\|_{\Lambda_{\boldsymbol{z}}}^2 + \frac{\mu_{\boldsymbol{s}}}{2}\left\|\boldsymbol{s} - \boldsymbol{x}^{(j+1)} - \eta_{\boldsymbol{s}}^{(j)}\right\|_{\Lambda_{\boldsymbol{s}}}^2, && (6.38)\\
\boldsymbol{z}^{(j+1)} &= \underset{\boldsymbol{z}}{arg\,min}\, \mathsf{R}(\boldsymbol{z}) + \frac{\mu_{\boldsymbol{z}}}{2}\left\|\boldsymbol{z} - \boldsymbol{C}\boldsymbol{s}^{(j+1)} - \eta_{\boldsymbol{z}}^{(j)}\right\|_{\Lambda_{\boldsymbol{z}}}^2, && (6.39)\\
\eta_{\boldsymbol{u}}^{(j+1)} &= \eta_{\boldsymbol{u}}^{(j)} - (\boldsymbol{u}^{(j+1)} - \boldsymbol{T}\boldsymbol{x}^{(j+1)}), && (6.40)\\
\eta_{\boldsymbol{v}}^{(j+1)} &= \eta_{\boldsymbol{v}}^{(j)} - (\boldsymbol{v}^{(j+1)} - \boldsymbol{A}\boldsymbol{u}^{(j+1)}), && (6.41)\\
\eta_{\boldsymbol{s}}^{(j+1)} &= \eta_{\boldsymbol{s}}^{(j)} - (\boldsymbol{s}^{(j+1)} - \boldsymbol{x}^{(j+1)}), && (6.42)\\
\eta_{\boldsymbol{z}}^{(j+1)} &= \eta_{\boldsymbol{z}}^{(j)} - (\boldsymbol{z}^{(j+1)} - \boldsymbol{C}\boldsymbol{s}^{(j+1)}), && (6.43)
\end{aligned}$$

The sub-problems (6.35) to (6.38) are all quadratic problems for which analytical solutions exist. However, (6.35) and (6.36) cannot be implemented explicitly due to the enormous sizes of the matrices involved. We employ the iterative CG-solver for these sub-problems.

Sub-problem (6.35) is an image-registration-type problem, which has the following analytical solution:

$$\boldsymbol{x}^{(j+1)} = \boldsymbol{H}^{-1}(\mu_{\boldsymbol{u}}\boldsymbol{T}^T\Lambda_{\boldsymbol{u}}(\boldsymbol{u}^{(j)} - \eta_{\boldsymbol{u}}^{(j)}) + \mu_{\boldsymbol{s}}\Lambda_{\boldsymbol{s}}(\boldsymbol{s}^{(j)} - \eta_{\boldsymbol{s}})), \tag{6.44}$$

and

$$\boldsymbol{H} = \mu_{\boldsymbol{u}}\boldsymbol{T}^T\boldsymbol{T} + \mu_{\boldsymbol{s}}\Lambda_{\boldsymbol{s}}, \tag{6.45}$$

is nonsingular because $\Lambda_{\boldsymbol{s}} = \boldsymbol{I}_N \succ 0$. Compared to the original problem (6.22), this sub-problem requires much less computation due to the absence of the system matrix $\boldsymbol{A}$ and the regularizer $\boldsymbol{R}$. We accelerate the CG-solver for (6.44) by using a suitable preconditioner for $\boldsymbol{H}$. Since $\boldsymbol{H}$ is much simpler than the Hessian of the original data term in (6.22), it is more amenable to preconditioning.

We now consider (6.36), which is a tomography problem with the following solution:

$$\boldsymbol{u}^{(j+1)} = \boldsymbol{M}^{-1}(\mu_{\boldsymbol{u}}\Lambda_{\boldsymbol{u}}(\boldsymbol{T}\boldsymbol{x}^{(j+1)} + \eta_{\boldsymbol{u}}^{(j)}) + \mu_{\boldsymbol{v}}\boldsymbol{A}^T(\boldsymbol{v}^{(j)} - \eta_{\boldsymbol{v}})), \tag{6.46}$$

and

$$\boldsymbol{M} = \mu_{\boldsymbol{v}}\boldsymbol{A}^T\boldsymbol{A} + \mu_{\boldsymbol{u}}\Lambda_{\boldsymbol{u}}, \tag{6.47}$$

is non-singular because $\Lambda_{\boldsymbol{u}} = \boldsymbol{I}_{NN_f} \succ 0$. We preconditioned this term to obtain faster convergence. Designing the preconditioners for $\boldsymbol{H}$ and $\boldsymbol{M}$ is discussed further in the following section. Further

more, this sub-problem can be further parallelized into N_f problems. Each parallelized problem can be efficiently solved by preconditioned CG or ordered-subsets type algorithms, which are less efficient for the original problem.

Sub-problems (6.37) to (6.39) can be solved much more easily compared to above two sub-problems. Sub-problem (6.37) has a simple analytical solution:

$$\boldsymbol{v}^{(j+1)} = \boldsymbol{D}^{-1}(\boldsymbol{W}\boldsymbol{y} + \mu_{\boldsymbol{v}}\Lambda_{\boldsymbol{v}}(\boldsymbol{A}\boldsymbol{u} + \eta_{\boldsymbol{v}})), \tag{6.48}$$

where $\boldsymbol{D} = (\boldsymbol{W} + \mu_{\boldsymbol{v}}\Lambda_{\boldsymbol{v}})$. Since $\boldsymbol{W}$ is diagonal, $\boldsymbol{D}$ can be inverted exactly leading to an exact solution.

We consider (6.38) which can be seen as an image denoising problem. This can be easily solved analytically:

$$\boldsymbol{s}^{(j+1)} = \boldsymbol{K}^{-1}(\mu_{\boldsymbol{z}}\boldsymbol{C}^T\Lambda_{\boldsymbol{z}}(\boldsymbol{z}^{(j)} - \eta_{\boldsymbol{z}}^{(j)}) + \mu_{\boldsymbol{s}}\Lambda_{\boldsymbol{s}}(\boldsymbol{x}^{(j)} + \eta_{\boldsymbol{s}})), \tag{6.49}$$

and

$$\boldsymbol{K} = \mu_{\boldsymbol{z}}\boldsymbol{C}^T\boldsymbol{C} + \mu_{\boldsymbol{s}}\Lambda_{\boldsymbol{s}}, \tag{6.50}$$

is non-singular because $\Lambda_{\boldsymbol{s}} = \boldsymbol{I}_N \succ 0$. Furthermore, we can see that $\boldsymbol{K}$ contains $\boldsymbol{C}^T\boldsymbol{C}$, which is shift-invariant for $\boldsymbol{C}$ with periodic end condition. So (6.38) is exactly solvable with Fourier transform if we use $\boldsymbol{C}$ with periodic end condition.

Finally, (6.39), which is also an image-denoising-type problem, can be solved easily with iterative algorithms or exactly solved for a variety of potential functions. Here, we consider one of the edge-preserving regularization using the Fair potential function:

$$\psi(t) = \delta^2[|t/\delta| - \log(1 + |t/\delta|)]. \tag{6.51}$$

For this regularizer, (6.39) separates into 1D minimization problems in terms of the component $\{z_k\}_{k=1}^K$ of $\boldsymbol{z}$:

$$z_k^{(j+1)} = \underset{z_k}{arg\,min}\, \beta\kappa_k\, \psi_k(z_k) + \frac{\mu_{\boldsymbol{z}}}{2}(z_k - \rho_k^{(j)})^2, \tag{6.52}$$

where $\rho_k^{(j)}$ is the kth component of $\rho^{(j)} \triangleq \boldsymbol{C}\boldsymbol{s}^{(j+1)} + \eta_{\boldsymbol{z}}^{(j)}$. By solving (6.52), we obtain the following exact solution:

$$z_k^{(j+1)} = \text{sign}\{\rho_k^{(j)}\}\frac{\zeta_k^{(j)} + \sqrt{(\zeta_k^{(j)})^2 + 4\delta\left|\rho_k^{(j)}\right|}}{2}, \tag{6.53}$$

where $\zeta_k^{(j)} \triangleq \left|\rho_k^{(j)}\right| - \delta - \delta\beta\kappa_k/\mu_z$.

Finally, we present our algorithm in Table 6.7 for solving (6.23).

Table 6.7: Splitting approach for motion-compensated image reconstruction in X-ray CT imaging

1. Select $\boldsymbol{x}^{(0)}$, $\mu_{\boldsymbol{u}}$, $\mu_{\boldsymbol{v}}$, $\mu_{\boldsymbol{z}} > 0$, $\mu_{\boldsymbol{s}} > 0$ and set j = 0
2. Set $\boldsymbol{u}^{(0)} = T(\underline{\alpha})\boldsymbol{x}^{(0)}$, $\boldsymbol{v}^{(0)} = \boldsymbol{A}\boldsymbol{u}^{(0)}$, $\boldsymbol{s}^{(0)} = \boldsymbol{x}^{(0)}$, $\boldsymbol{z}^{(0)} = \boldsymbol{C}\boldsymbol{s}^{(0)}$, and $\eta_{\boldsymbol{u}}^{(0)} = \eta_{\boldsymbol{v}}^{(0)} = \eta_{\boldsymbol{s}}^{(0)} = \eta_{\boldsymbol{z}}^{(0)} = \epsilon$

Repeat:

3. Obtain $\boldsymbol{x}^{(j+1)}$ applying PCG to (6.35)
4. Obtain $\boldsymbol{u}^{(j+1)}$ applying PCG to (6.36)
5. Obtain $\boldsymbol{v}^{(j+1)}$ using (6.48)
6. Obtain $\boldsymbol{s}^{(j+1)}$ using (6.49)
7. Obtain $\boldsymbol{z}^{(j+1)}$ using (6.53)
8. $\eta_{\boldsymbol{u}}{}^{(j+1)} = \eta_{\boldsymbol{u}}{}^{(j)} - (\boldsymbol{u}^{(j+1)} - \boldsymbol{T}(\underline{\alpha})\boldsymbol{x}^{(j+1)})$
9. $\eta_{\boldsymbol{v}}{}^{(j+1)} = \eta_{\boldsymbol{v}}{}^{(j)} - (\boldsymbol{v}^{(j+1)} - \boldsymbol{A}\boldsymbol{u}^{(j+1)})$
10. $\eta_{\boldsymbol{s}}{}^{(j+1)} = \eta_{\boldsymbol{s}}{}^{(j)} - (\boldsymbol{s}^{(j+1)} - \boldsymbol{x}^{(j+1)})$
11. $\eta_{\boldsymbol{z}}{}^{(j+1)} = \eta_{\boldsymbol{z}}{}^{(j)} - (\boldsymbol{z}^{(j+1)} - \boldsymbol{C}\boldsymbol{s}^{(j+1)})$
12. Set $j = j + 1$

Until stopping criterion is satisfied

6.2.2.4 Preconditioner designs

Using an efficient preconditioner will help each sub-problem converge with fewer iterations. Reducing these sub-iteration numbers while obtaining the solution with compatible quality leads to faster convergence in time. Therefore, designing a decent preconditioner for each sub-problem is very important for the speed of the proposed method.

First, consider the sub-problem (6.44). In general, using a diagonal preconditioner does not improve the convergence rate much [41], and since $\boldsymbol{H}$ is shift-variant, using a circulant preconditioner would not work well neither. Since the sub-problem (6.44) is computationally much inexpensive compared to the sub-problem (6.46), the preconditioner for $\boldsymbol{H}$ does not critically affect the overall convergence speed. Therefore, we used a simple a diagonal preconditioner for this sub-problem and leave the design of more efficient preconditioner as future work. By extracting the diagonal elements of $\boldsymbol{H}$, we use its inverse as the preconditioner.

Next, consider the sub-problem (6.46). This step is computationally expensive compared to other steps, so solving this sub-problem efficiently is critical to make our method practical. Here, we propose new preconditioners and compare them with other well-known preconditioners.

Circulant preconditioner Circulant preconditioner is widely used for tomographic reconstruction problems [29, 105]. When the hessian of the cost function can be approximated as shift-invariant, the circulant preconditioner can accelerate the convergence efficiently.

We can see that $\boldsymbol{M}$ contains $\boldsymbol{A}^T\boldsymbol{A}$, which can be rewritten as:

$$\boldsymbol{A}^T\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_1^T\boldsymbol{A}_1 & & & \\ & \boldsymbol{A}_2^T\boldsymbol{A}_2 & & \\ & & \ddots & \\ & & & \boldsymbol{A}_{N_f}^T\boldsymbol{A}_{N_f} \end{bmatrix}, \tag{6.54}$$

where $\boldsymbol{A}_m^T\boldsymbol{A}_m$ is "nearly" shift-invariant. So we can precondition the problem using suitable cone filters [41]. First, we can rewrite $\boldsymbol{M}$ as follows:

$$\begin{aligned} \boldsymbol{M} &= \begin{bmatrix} \boldsymbol{M}_1 & & \\ & \ddots & \\ & & \boldsymbol{M}_{N_f} \end{bmatrix} \\ &= \begin{bmatrix} \mu_{\boldsymbol{v}}\boldsymbol{A}_1^T\boldsymbol{A}_1 + \mu_{\boldsymbol{u}}\boldsymbol{I}_N & & \\ & \ddots & \\ & & \mu_{\boldsymbol{v}}\boldsymbol{A}_{N_f}^T\boldsymbol{A}_{N_f} + \mu_{\boldsymbol{u}}\boldsymbol{I}_N \end{bmatrix}. \end{aligned} \tag{6.55}$$

For each $\boldsymbol{M}_m$, we can construct a circulant matrix $\bar{\boldsymbol{M}}_m$ from the central column of $\boldsymbol{M}_m$:

$$\bar{\boldsymbol{M}}_m = \mathrm{circ}(\boldsymbol{M}_m \mathsf{e}_c), \tag{6.56}$$

and use its inverse, $\bar{\boldsymbol{M}}_m^{-1}$, as the preconditioner, where e_c is a standard basis vector of $\mathbb{R}^N$ corresponding to the center pixel of the image and circ(ρ) represents the construction of a circulant matrix from a vector ρ. Therefore, the circulant preconditioner for the problem is given by:

$$\bar{\boldsymbol{M}}^{-1} = \begin{bmatrix} \bar{\boldsymbol{M}}_1^{-1} & & \\ & \ddots & \\ & & \bar{\boldsymbol{M}}_{N_f}^{-1} \end{bmatrix} \tag{6.57}$$

Shift-variant preconditioner In above, we assumed that $\boldsymbol{A}_m^T \boldsymbol{A}_m$ is "nearly" shift-invariant. However, this assumption may not be reasonable especially when N_f is large. Here, we introduce another preconditioner for $\boldsymbol{M}$, which was suggested by [41]. We can approximate $\boldsymbol{A}_m^T \boldsymbol{A}_m$ as follows:

$$\boldsymbol{A}_m^T \boldsymbol{A}_m = \boldsymbol{A}_0^T \boldsymbol{W}_m \boldsymbol{A}_0 \approx \boldsymbol{K}_m \boldsymbol{A}_0^T \boldsymbol{A}_0 \boldsymbol{K}_m, \tag{6.58}$$

where $\boldsymbol{A}_0 = [\boldsymbol{A}_1^T, \cdots, \boldsymbol{A}_{N_f}^T]^T$, $\boldsymbol{W}_m$ is a diagonal matrix which only selects the measurements obtained from $\boldsymbol{A}_m$, $\boldsymbol{K}_m = \boldsymbol{D}[\kappa_{m,j}]$ and

$$\kappa_{m,j} = \sqrt{\frac{\sum_i a_{ij}^2 [\boldsymbol{W}_m]_{ii}}{\sum_i a_{ij}^2}} \approx \sqrt{\frac{\sum_i a_{ij} [\boldsymbol{W}_m]_{ii}}{\sum_i a_{ij}}}, \tag{6.59}$$

In (6.59), we approximated once again with a_{ij} instead of a_{ij}^2 for computational convenience. We can obtain approximate $\boldsymbol{M}$ as follows:

$$\begin{aligned} \boldsymbol{M} &= \begin{bmatrix} \mu_{\boldsymbol{v}} \boldsymbol{A}_1^T \boldsymbol{A}_1 + \mu_{\boldsymbol{u}} \boldsymbol{I}_N & & \\ & \ddots & \\ & & \mu_{\boldsymbol{v}} \boldsymbol{A}_{N_f}^T \boldsymbol{A}_{N_f} + \mu_{\boldsymbol{u}} \boldsymbol{I}_N \end{bmatrix} \\ &\approx \begin{bmatrix} \mu_{\boldsymbol{v}} \boldsymbol{K}_1 \boldsymbol{A}_0^T \boldsymbol{A}_0 \boldsymbol{K}_1 + \mu_{\boldsymbol{u}} \boldsymbol{I}_N & & \\ & \ddots & \\ & & \mu_{\boldsymbol{v}} \boldsymbol{K}_{N_f} \boldsymbol{A}_0^T \boldsymbol{A}_0 \boldsymbol{K}_{N_f} + \mu_{\boldsymbol{u}} \boldsymbol{I}_N \end{bmatrix}. \end{aligned} \tag{6.60}$$

For each approximated $\boldsymbol{M}_m$, we get:

$$\boldsymbol{M}_m \approx \mu_{\boldsymbol{v}} \boldsymbol{K}_m \boldsymbol{A}_0^T \boldsymbol{A}_0 \boldsymbol{K}_m + \mu_{\boldsymbol{u}} \boldsymbol{I}_N = \boldsymbol{K}_m \boldsymbol{J}_m \boldsymbol{K}_m, \tag{6.61}$$

where $\boldsymbol{J}_m = \mu_{\boldsymbol{v}} \boldsymbol{A}_0^T \boldsymbol{A}_0 + \mu_{\boldsymbol{u}} \boldsymbol{K}_m^{-1} \boldsymbol{I}_N \boldsymbol{K}_m^{-1}$. We get the following approximated preconditioner for $\boldsymbol{M}$:

$$\boldsymbol{M}^{-1} \approx \begin{bmatrix} \boldsymbol{K}_1^{-1} \boldsymbol{J}_1^{-1} \boldsymbol{K}_1^{-1} & & \\ & \ddots & \\ & & \boldsymbol{K}_{N_f}^{-1} \boldsymbol{J}_{N_f}^{-1} \boldsymbol{K}_{N_f}^{-1} \end{bmatrix} \tag{6.62}$$

By using the approximations from [41], we can obtain the following shift-variant preconditioner for $\boldsymbol{M}_m$:

$$\begin{aligned} \boldsymbol{M}_M^{-1} &= \begin{bmatrix} \boldsymbol{M}_{M,1} & & \\ & \ddots & \\ & & \boldsymbol{M}_{M,N_f} \end{bmatrix} && (6.63) \\ \boldsymbol{M}_{M,m} &= \boldsymbol{K}_m^{-1} \boldsymbol{S}_m{}' \boldsymbol{S}_m \boldsymbol{K}_m^{-1}, && (6.64) \\ \boldsymbol{S}_m &\triangleq \sum_{k=1}^{l} \Omega_{m,k}^{-1/2} \boldsymbol{Q} \boldsymbol{D}_{\lambda_{m,k}}, && (6.65) \end{aligned}$$

where $\boldsymbol{Q}$ is the orthonormal version of the DFT operator, $\Omega_{m,k}$ is the k-sampled 2D DFT coefficients of $\boldsymbol{M}_m$, $\boldsymbol{D}_{\lambda_{m,k}}$ is a diagonal matrix with elements $\lambda_{m,k}(\eta_j)$, $\lambda_{m,k}$ is an interpolation factor, and η_j is an effective regularization parameter for the jth voxel.

Ad hoc design Since $\boldsymbol{J}_m$ is shift-variant, we cannot directly use circulant approximation, and thus we have used a shift-variant preconditioner. However, a shift-variant preconditioner requires more calculations than a circulant preconditioner. Instead of approximating the inverse of $\boldsymbol{J}_m$ with a shift-variant preconditioner, we use the following simple preconditioner:

$$\boldsymbol{M}^{-1} \approx \begin{bmatrix} \boldsymbol{K}_1^{-1} \bar{\boldsymbol{J}}_1^{-1} \boldsymbol{K}_1^{-1} & & \\ & \ddots & \\ & & \boldsymbol{K}_{N_f}^{-1} \bar{\boldsymbol{J}}_{N_f}^{-1} \boldsymbol{K}_{N_f}^{-1} \end{bmatrix}, \tag{6.66}$$

where $\bar{\boldsymbol{J}}_m = \operatorname{circ}\left(\left(\mu_{\boldsymbol{v}} \boldsymbol{A}_0^T \boldsymbol{A}_0 + \mu_{\boldsymbol{u}}^* \boldsymbol{I}_N \right) \mathsf{e}_c \right)$ with some new value of $\mu_{\boldsymbol{u}}^*$. Obtaining the optimal value for $\mu_{\boldsymbol{u}}^*$ may not be straightforward. We selected the value empirically.

Preconditioning with new weighting matrix So far, we have used the identity matrix for the symmetric weighting matrix, $\Lambda_{\boldsymbol{u}}$. However, we propose to use the following as the weighting matrix:

$$\Lambda_{\boldsymbol{u}} = \begin{bmatrix} \boldsymbol{K}_1^2 & & \\ & \ddots & \\ & & \boldsymbol{K}_{N_f}^2 \end{bmatrix}. \tag{6.67}$$

With this new weighting matrix, (6.47) can be approximated as:

$$\boldsymbol{M} \approx \begin{bmatrix} \boldsymbol{K}_1 \left(\mu_{\boldsymbol{v}} \boldsymbol{A}_0^T \boldsymbol{A}_0 + \mu_{\boldsymbol{u}} \boldsymbol{I}_N\right) \boldsymbol{K}_1 & & \\ & \ddots & \\ & & \boldsymbol{K}_{N_f} \left(\mu_{\boldsymbol{v}} \boldsymbol{A}_0^T \boldsymbol{A}_0 + \mu_{\boldsymbol{u}} \boldsymbol{I}_N\right) \boldsymbol{K}_{N_f} \end{bmatrix} \tag{6.68}$$

We can see that $\mu_{\boldsymbol{v}} \boldsymbol{A}_0^T \boldsymbol{A}_0 + \mu_{\boldsymbol{u}} \boldsymbol{I}_N$ is close to shift-invariant, and we can use circulant approximation to obtain the inverse of this new $\boldsymbol{M}$.

6.2.2.5 Parameter selection

The parameters $\mu_{\boldsymbol{u}}$, $\mu_{\boldsymbol{v}}$, $\mu_{\boldsymbol{z}}$, and $\mu_{\boldsymbol{s}}$ do not affect the solution of (6.22), but determine the convergence speed of the proposed splitting method [103]. Since steps 3 and 4 are inexact, the convergence speed of the proposed method is determined by how efficiently equations (6.35) and (6.36) are solved. Thus, selecting the parameters can become very critical to the performance of the proposed method. Not only their absolute values but also their relative values affect the convergence speed. Thus we first determine the value of $\mu_{\boldsymbol{v}}$, which can be obtained from (6.48), and then select the values of $\mu_{\boldsymbol{u}}$, $\mu_{\boldsymbol{z}}$, and $\mu_{\boldsymbol{s}}$ with fixed $\mu_{\boldsymbol{v}}$ value.

6.2.2.6 Selecting $\mu_{\boldsymbol{v}}$

Observing (6.48), we selected $\mu_{\boldsymbol{v}} = median\{w_i\}$ to avoid outliers in $\boldsymbol{W}$, which yields a well-conditioned $\boldsymbol{D}$ (with $\kappa\{\boldsymbol{D}\} \approx 10$ for our simulation). To see how robust the proposed method is to the choice of $\mu_{\boldsymbol{v}}$, we may perturb the value of $\mu_{\boldsymbol{v}}$ and see its effect on the convergence speed. It seems that the method is slightly more robust to the choice of $\mu_{\boldsymbol{v}}$ than $\mu_{\boldsymbol{u}}$, thus we select its value first.

6.2.2.7 Selecting $\mu_{\boldsymbol{u}}$

Since $\mu_{\boldsymbol{u}}$ and $\mu_{\boldsymbol{v}}$ balances $\mu_{\boldsymbol{v}} \boldsymbol{A}^T \boldsymbol{A}$ and $\mu_{\boldsymbol{u}} \boldsymbol{I}_N$ in $\boldsymbol{M}$, the condition number of $\boldsymbol{M}$ is determined by the ratio of these two. We fixed $\mu_{\boldsymbol{v}}$ as the value selected above, and chose $\mu_{\boldsymbol{u}}$ to obtain a reasonable balance between the convergence speed and the condition number. Too large value of $\mu_{\boldsymbol{u}}$ results in slow convergence. On the other extreme, setting $\mu_{\boldsymbol{u}}$ too small will lead to a poorly conditioned $\boldsymbol{M}$, which is not favorable. Based on our experiments, for $\boldsymbol{A}$ implemented using the separable footprints (SF) projector [79], we found that using the ratio $\mu_{\boldsymbol{u}}/\mu_{\boldsymbol{v}} \approx 10^3$ gives good overall convergence speeds for the proposed method. With this value, the condition number of $\bar{\boldsymbol{M}}$ is approximately 1.6.

6.2.2.8 Selecting μ_s

The parameter μ_s appears both (6.35) and (6.38). With fixed μ_u, we determine the value of μ_s based on (6.35). The ratio μ_s/μ_u determines the characteristics of $\boldsymbol{H}$, (6.45). As mentioned before, selecting the parameters based on the condition number is not straightforward. Also, estimating the condition number of T^TT is not straightforward. Thus selecting parameters based on the condition number of $\boldsymbol{H}$ is not promising. We will use the power iteration to obtain the maximum eigenvalue of T^TT and then use it to determine the parameter, μ_s. We did simulations with various values of μ_s to see how the value of μ_s affect the convergence speed of the method, and determined the value empirically.

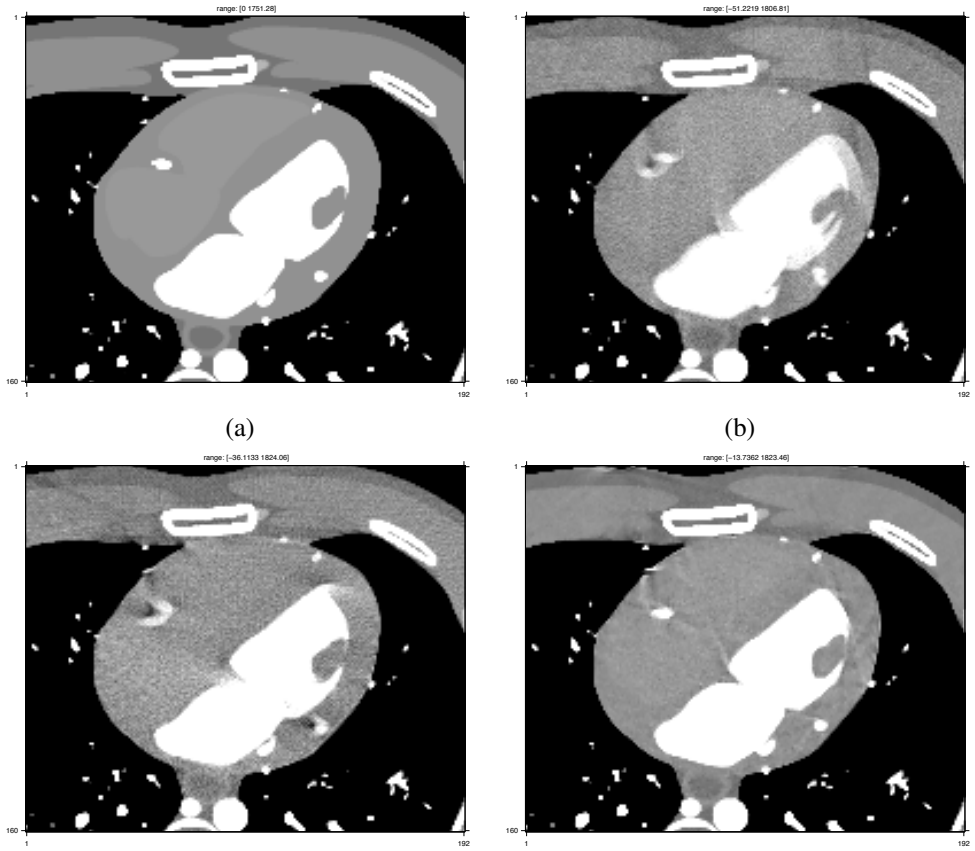
6.2.2.9 Selecting μ_z

The parameter μ_z appears (6.38). We fixed μ_s as the value selected above, and empirically chose μ_z to obtain a reasonable balance between the convergence speed and the condition number. Too small value of μ_z results in slow convergence. On the other extreme, setting μ_z too large will lead to a poorly conditioned $\boldsymbol{K}$, which is not favorable.

6.2.3 Simulation results

The proposed algorithm was investigated on a 2D CT image reconstruction problem with cardiac motion for simulated data. We simulated a 3rd-generation fan-beam CT system using the separable footprint projector [79]. The simulated system has 888 channels per view spaced 1.0239 mm apart, and 984 evenly spaced view angles over 360°. The image was reconstructed to a 512 $\times$ 512 grid of 0.9766 mm pixels. We generated seven frames of the XCAT phantom for a heart rate of 75 bpm. The motion between the frames was estimated directly from XCAT images using nonrigid image registration. Estimating motion parameters from true images is unrealistic, but our focus is not on obtaining reasonable motion estimates. We only focus on the image reconstruction part of MCIR. For the regularizer, we used a Fair potential function to provide edge-preservation and a certainty-based penalty to obtain more uniform resolution. The sinogram was generated with Poisson noise, and the weights in the data-fit term in (6.22) were chosen as $w_i = \exp(-[\boldsymbol{A}\boldsymbol{x}]_i)$. We selected the regularization parameter β such that the target PSF has a full-width at half-maximum (FWHM) of approximately 1.3 mm.

For comparison, we used the (nonlinear) conjugate gradient algorithm to solve the original problem (6.22). To analyze the convergence speed of the proposed method we computed the root mean squared (RMS) difference between the image estimate at the nth iteration, $\boldsymbol{x}^{(n)}$, with the "fully" converged solution, $\boldsymbol{x}^{\infty}$. For the Fair potential, the original MCIR problem is strictly convex and thus has a unique minimizer, $\boldsymbol{x}^{\infty}$. We numerically approximated $\boldsymbol{x}^{\infty}$ as the mean of the



(a) (b)

(c) (d)

Figure 6.12: Images in the ROI of (a) XCAT phantom, (b) Full-scan FBP reconstruction (c) Short-scan FBP reconstruction (also the initial guess $\boldsymbol{x}^{(0)}$), (d) Converged Image $\boldsymbol{x}^{(\infty)}$.

images reconstructed (assuming convergence) by running 1000 iterations of CG and 700 iterations of the proposed method with (10,10) sub-iterations.

Fig. 6.12 illustrates that the conventional filtered backprojection (FBP) methods, both full-scan and short-scan, gives reconstructed images with severe motion artifacts but the motion-compensated image, on the contrary, contains much less motion artifacts. Some residual motion artifacts still exist due to imperfect motion estimates even though they were obtained directly from the true XCAT images.

Fig. 6.13 illustrates that the proposed method converges much faster in iterations compared to the conventional CG method when we use enough sub-iterations with obvious computation overhead. This result suggests that if we have a proper preconditioner for each sub-problem, we can still obtain fast convergence. We also investigated different options for the proposed method summarized in Fig. 6.13. Using a preconditioner for sub-problems helped reduce the number of sub-iterations while achieving fast convergence speed. Use of ordered-subsets methods was also investigated. Both conventional OS and OS + momentum showed fast convergence speed initially,

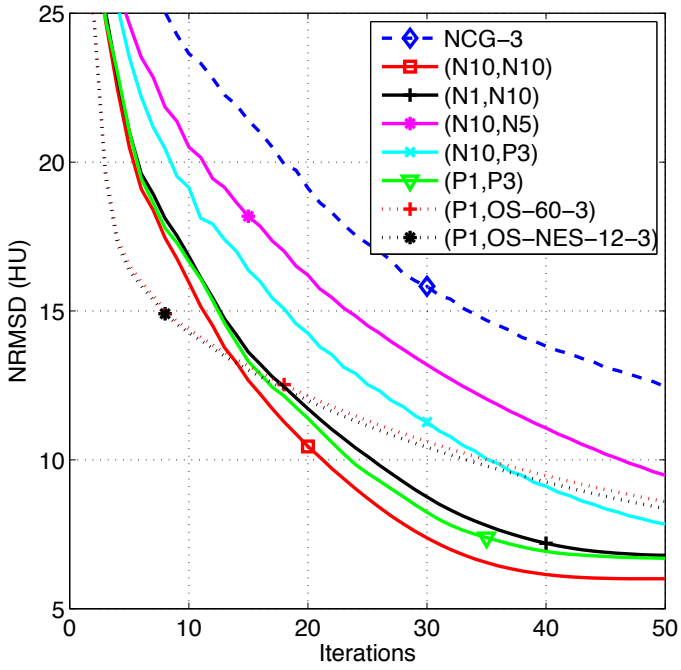


Figure 6.13: Plot of RMSD versus iteration for various settings of the proposed method compared to the conventional CG method. For the proposed method, (Nα,Pγ) indicates α iterations for sub-problem (6.35) without preconditioner for $\boldsymbol{H}$ and γ iterations for sub-problem (6.36) with a preconditioner for $\boldsymbol{M}$. OS-60-η and OS-NES-12-η indicate ordered subsets method with 60 subsets and "OS + momentum method" with 12 subsets, respectively, and η is the number of iterations.

but slowed down after certain number of iterations.

We have compared the convergence speed of our method with various preconditioners in Fig. 6.14. All the conditions, except for preconditioner, were fixed. We used a diagonal preconditioner for the registration part and different preconditioners for the tomography part. We can observe that the circulant preconditioner with new weighting (6.67) gives much faster convergence speed compared to other preconditioners.

In Fig. 6.15, we compare the proposed method with circulant preconditioner and new weighting (6.67) to the conventional CG method. In Fig. 6.14, the parameters were not chosen to compete with the conventional CG, so they were sub-optimal. For the purpose of illustration, we selected 5 and 2 sub-iterations for (6.35) and (6.36), respectively. The proposed method shows much faster convergence speed compared to CG method. While the proposed method as implemented in MATLAB provides marginal improvement in convergence speed over CG, we believe its ability to parallelize some of the update steps can further improve its efficiency.

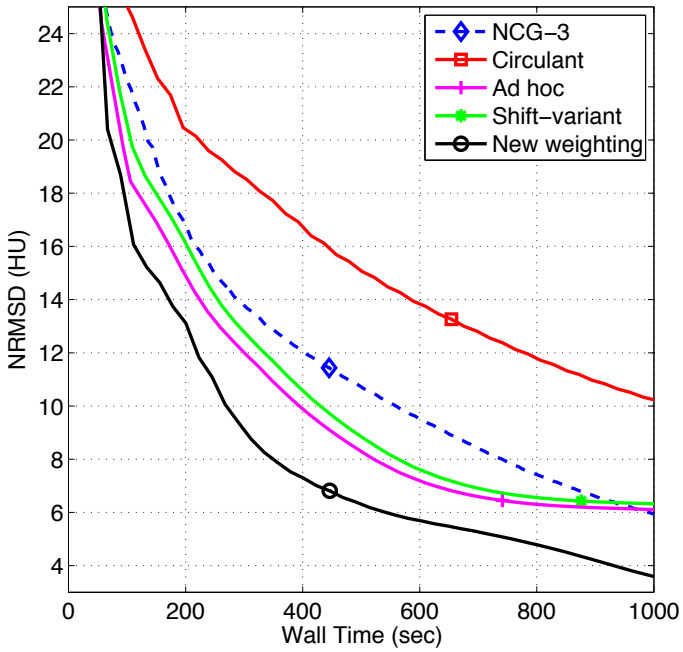


Figure 6.14: Compare the convergence speed of different preconditioners

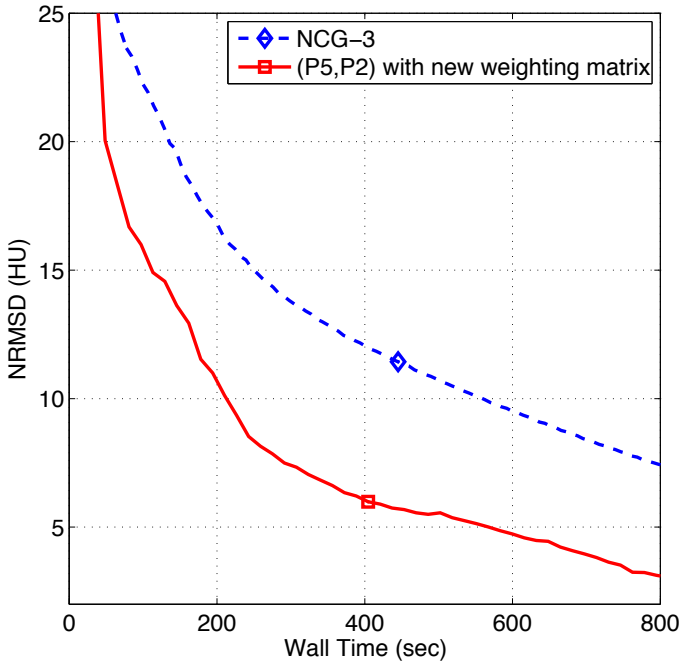


Figure 6.15: Plot of RMSD versus wall time for the proposed method compared to the conventional CG method with 3 line-search iterations. A diagonal preconditioner and a circulant preconditioner were used for sub-problems (6.35) and (6.36) respectively.

6.2.4 Discussion

We applied a variable splitting approach to the motion-compensated image reconstruction problem. The proposed method has faster convergence speed than the conjugate gradient method, and

offers the potential for parallelizability and preconditioning of sub-problems. Some of the sub-problems can be solved simultaneously or further divided into smaller problems. By using more sophisticated preconditioners for the sub-problems, the proposed method can be further improved. In this study, we focussed on the image reconstruction part of MCIR, but our method also can be extended to the joint estimation framework. Our future work will focus on improving the convergence speed of the proposed method and on applying it to 3-D CT.

6.3 Accelerating the joint motion estimation and image reconstruction using a variable splitting approach

Assuming a known motion may not be reasonable since obtaining motion parameters is not trivial. Inaccurate motion estimates lead to residual artifacts, and thus relying on initial motion estimates may results in unacceptable reconstructed images. Joint estimation framework has potential to solve this issue. In this section section, we focus on the formulation of joint motion estimation and image reconstruction framework for MCIR problem and its acceleration using a variable splitting approach. Acceleration of joint MCIR was not evaluated yet, thus we only provide the idea and a sinogram-based coronary motion estimation method for estimating initial motion parameters.

6.3.1 Joint motion estimation and image reconstruction for CT

As mentioned previously, the quality of the reconstructed images of MCIR methods depend heavily of the accuracy of estimated motion parameters. To compensate for imperfect initial motion estimations, joint estimation for both motion parameters and reconstructed images are often introduced [122, 123]. Joint estimation of motion parameters and reconstructed images has potential to provide better estimators at the expense of increased computation. In this section, we apply a variable splitting technique similar to that of previous section.

For motion-compensated CT image reconstruction, we consider following joint registration and reconstruction formulation:

$$(\hat{\boldsymbol{x}}, \hat{\underline{\alpha}}) = \underset{\boldsymbol{x}, \underline{\alpha}}{arg\,min} \left\{ \Psi(\boldsymbol{x}, \underline{\alpha}) \triangleq \text{Ł}(\boldsymbol{x}, \underline{\alpha}) + \mathsf{R}_1(\boldsymbol{x}) + \mathsf{R}_2(\underline{\alpha}) \right\}, \tag{6.69}$$

$$\text{Ł}(\boldsymbol{x}, \underline{\alpha}) = \frac{1}{2} \| \boldsymbol{y} - \boldsymbol{A}\boldsymbol{T}(\underline{\alpha})\boldsymbol{x} \|_{\boldsymbol{W}}^2, \quad \mathsf{R}_2(\underline{\alpha}) = \sum_{m=1}^{N_f} \beta_{2,m} \mathsf{R}_\alpha(\underline{\alpha}_m),$$

$$\boldsymbol{A} = \text{diag}\{\boldsymbol{A}_1, \cdots, \boldsymbol{A}_{N_f}\}, \quad \boldsymbol{T}(\underline{\alpha}) = [\boldsymbol{T}(\underline{\alpha}_1)^T \dots \boldsymbol{T}(\underline{\alpha}_{N_f})^T]^T, \quad \underline{\alpha} = (\underline{\alpha}_1, \cdots, \underline{\alpha}_{N_f}),$$

where R_1 is the regularizer for the reconstructed image defined in (2.17), R_2 is the regularizer for

the motion parameters (see below), $\beta_{2,m}$ is a regularization coefficient, $\boldsymbol{T}(\cdot)$ is the warp matrix, and $\underline{\alpha}_m$ is the motion parameter for mth frame. For motion parameters of each frame, $\underline{\alpha}_m$, we use piecewise quadratic regularizer suggested in [28]:

$$\begin{aligned}\mathsf{R}_2(\alpha) = \sum_{q\in\{x,y,z\}}\sum_{i,j,k}[p(\alpha^q_{i+1,j,k}-\alpha^q_{i,j,k};\zeta_1^{q,x},\zeta_2^{q,x}) + p(\alpha^q_{i,j+1,k}-\alpha^q_{i,j,k};\zeta_1^{q,y},\zeta_2^{q,y}) \\ + p(\alpha^q_{i,j,k+1}-\alpha^q_{i,j,k};\zeta_1^{q,z},\zeta_2^{q,z})],\end{aligned} \tag{6.70}$$

where

$$p(t;\zeta_1,\zeta_2) = \begin{cases} \frac{1}{2}(t-\zeta_1)^2, & t<\zeta_1 \\ 0, & \zeta_1\le t\le\zeta_2 \\ \frac{1}{2}(t-\zeta_2)^2. & \zeta_2<t \end{cases}$$

6.3.2 Proposed method for accelerated joint estimation

6.3.2.1 Equivalent constraint optimization problem

We introduce auxiliary constraint variables $\boldsymbol{u}$ and $\boldsymbol{v}$, and write (6.69) as the following equivalent constrained problem:

$$\begin{aligned}&\underset{\boldsymbol{x},\underline{\alpha},\boldsymbol{u},\boldsymbol{v},\boldsymbol{z}}{arg\,min}\,\Psi(\boldsymbol{x},\boldsymbol{u},\boldsymbol{v},\boldsymbol{z}) = \frac{1}{2}\|\boldsymbol{y}-\boldsymbol{v}\|^2_{\boldsymbol{W}} + \mathsf{R}_1(\boldsymbol{z}) + \mathsf{R}_2(\underline{\alpha}), \\ &s.t. \quad \boldsymbol{u}=\boldsymbol{T}(\underline{\alpha})\boldsymbol{x}, \quad \boldsymbol{v}=\boldsymbol{A}\boldsymbol{u}, \quad \boldsymbol{z}=\boldsymbol{x},\end{aligned} \tag{6.71}$$

where $\boldsymbol{u}\in\Re^{NN_f}$ separates the system matrix from the warp matrix, $\boldsymbol{v}\in\Re^M$ separates the effect of the weighting matrix, $\boldsymbol{W}$, on $\boldsymbol{A}\boldsymbol{x}$, and $\boldsymbol{z}\in\Re^N$ detaches the warp matrix from the regularizer.

6.3.2.2 Method of multipliers

To solve (6.71), we use the framework of the method of multipliers and construct an augmented Lagrangian function as follows:

$$\begin{aligned}L(\boldsymbol{x},\boldsymbol{u},\boldsymbol{v},\boldsymbol{z}) \;&\triangleq\; \frac{1}{2}\|\boldsymbol{y}-\boldsymbol{v}\|^2_{\boldsymbol{W}} + \mathsf{R}_1(\boldsymbol{z}) + \mathsf{R}_2(\underline{\alpha}) \\ &+\; \gamma_{\boldsymbol{u}}^T(\boldsymbol{u}-\boldsymbol{T}(\underline{\alpha})\boldsymbol{x}) + \frac{\mu_{\boldsymbol{u}}}{2}\|\boldsymbol{u}-\boldsymbol{T}(\underline{\alpha})\boldsymbol{x}\|^2_{\Lambda_{\boldsymbol{u}}} \\ &+\; \gamma_{\boldsymbol{v}}^T(\boldsymbol{v}-\boldsymbol{A}\boldsymbol{u}) + \frac{\mu_{\boldsymbol{v}}}{2}\|\boldsymbol{v}-\boldsymbol{A}\boldsymbol{u}\|^2_{\Lambda_{\boldsymbol{v}}} + \gamma_{\boldsymbol{z}}^T(\boldsymbol{z}-\boldsymbol{x}) + \frac{\mu_{\boldsymbol{z}}}{2}\|\boldsymbol{z}-\boldsymbol{x}\|^2_{\Lambda_{\boldsymbol{z}}},\end{aligned}$$

where $\gamma_{\boldsymbol{u}}\in\Re^{NN_f}$, $\gamma_{\boldsymbol{v}}\in\Re^M$ and $\gamma_{\boldsymbol{z}}\in\Re^N$ are the Lagrange multipliers, $\mu_{\boldsymbol{u}}>0$, $\mu_{\boldsymbol{v}}>0$, and $\mu_{\boldsymbol{z}}>0$ are the AL penalty parameters, and $\Lambda_{\boldsymbol{u}}\succ 0$, $\Lambda_{\boldsymbol{v}}\succ 0$, and $\Lambda_{\boldsymbol{z}}\succ 0$ are symmetric weighting

matrixes. The multiplier term can be absorbed into the quadratic penalty term, and by ignoring the irrelevant constants we obtain:

$$\begin{aligned} L(\boldsymbol{x},\boldsymbol{u},\boldsymbol{v},\boldsymbol{z}) &\triangleq \frac{1}{2}\|\boldsymbol{y}-\boldsymbol{v}\|^2_{\boldsymbol{W}} + \mathsf{R}_1(\boldsymbol{z}) + \mathsf{R}_2(\underline{\alpha}) \\ &+ \frac{\mu_{\boldsymbol{u}}}{2}\|\boldsymbol{u}-\boldsymbol{T}(\underline{\alpha})\boldsymbol{x}-\eta_{\boldsymbol{u}}\|^2_{\Lambda_{\boldsymbol{u}}} + \frac{\mu_{\boldsymbol{v}}}{2}\|\boldsymbol{v}-\boldsymbol{A}\boldsymbol{u}-\eta_{\boldsymbol{v}}\|^2_{\Lambda_{\boldsymbol{v}}} + \frac{\mu_{\boldsymbol{z}}}{2}\|\boldsymbol{z}-\boldsymbol{x}-\eta_{\boldsymbol{z}}\|^2_{\Lambda_{\boldsymbol{z}}}, \end{aligned} \tag{6.72}$$

where $\eta_{\boldsymbol{u}} = -\frac{1}{\mu_{\boldsymbol{u}}}\Lambda_{\boldsymbol{u}}^{-1}\gamma_{\boldsymbol{u}}$, $\eta_{\boldsymbol{v}} = -\frac{1}{\mu_{\boldsymbol{v}}}\Lambda_{\boldsymbol{v}}^{-1}\gamma_{\boldsymbol{v}}$, and $\eta_{\boldsymbol{z}} = -\frac{1}{\mu_{\boldsymbol{z}}}\Lambda_{\boldsymbol{z}}^{-1}\gamma_{\boldsymbol{z}}$. We use $\Lambda_{\boldsymbol{u}} = \boldsymbol{I}_{NN_f}$, $\Lambda_{\boldsymbol{v}} = \boldsymbol{I}_M$, and $\Lambda_{\boldsymbol{z}} = \boldsymbol{I}_N$, which is a standard approach.

The classical AL scheme for solving the problem (6.71) is to alternate between a joint-minimization step and an update step:

$$(\boldsymbol{x}^{(j+1)},\underline{\alpha}^{(j+1)},\boldsymbol{u}^{(j+1)},\boldsymbol{v}^{(j+1)},\boldsymbol{z}^{(j+1)}) = \underset{\boldsymbol{x},\underline{\alpha},\boldsymbol{u},\boldsymbol{v},\boldsymbol{z}}{arg\,min}\, L(\boldsymbol{x},\underline{\alpha},\boldsymbol{u},\boldsymbol{v},\boldsymbol{z},\eta_{\boldsymbol{u}}^{(j)},\eta_{\boldsymbol{v}}^{(j)},\eta_{\boldsymbol{z}}^{(j)}), \tag{6.73}$$

$$\eta_{\boldsymbol{u}}^{(j+1)} = \eta_{\boldsymbol{u}}^{(j)} - (\boldsymbol{u}^{(j+1)} - \boldsymbol{T}(\underline{\alpha})\boldsymbol{x}^{(j+1)}), \tag{6.74}$$

$$\eta_{\boldsymbol{v}}^{(j+1)} = \eta_{\boldsymbol{v}}^{(j)} - (\boldsymbol{v}^{(j+1)} - \boldsymbol{A}\boldsymbol{u}^{(j+1)}), \tag{6.75}$$

$$\eta_{\boldsymbol{z}}^{(j+1)} = \eta_{\boldsymbol{z}}^{(j)} - (\boldsymbol{z}^{(j+1)} - \boldsymbol{x}^{(j+1)}), \tag{6.76}$$

6.3.2.3 Alternating direction minimization

We replace the joint-minimization step (6.73) by alternating direction optimization as follows:

$$\boldsymbol{x}^{(j+1)} = \underset{\boldsymbol{x}}{arg\,min}\, L(\boldsymbol{x},\underline{\alpha}^{(j)},\boldsymbol{u}^{(j)},\boldsymbol{v}^{(j)},\boldsymbol{z}^{(j)},\eta_{\boldsymbol{u}}^{(j)},\eta_{\boldsymbol{v}}^{(j)},\eta_{\boldsymbol{z}}^{(j)}), \tag{6.77}$$

$$\underline{\alpha}^{(j+1)} = \underset{\underline{\alpha}}{arg\,min}\, L(\boldsymbol{x}^{(j+1)},\underline{\alpha},\boldsymbol{u}^{(j)},\boldsymbol{v}^{(j)},\boldsymbol{z}^{(j)},\eta_{\boldsymbol{u}}^{(j)},\eta_{\boldsymbol{v}}^{(j)},\eta_{\boldsymbol{z}}^{(j)}), \tag{6.78}$$

$$\boldsymbol{u}^{(j+1)} = \underset{\boldsymbol{u}}{arg\,min}\, L(\boldsymbol{x}^{(j+1)},\underline{\alpha}^{(j+1)},\boldsymbol{u},\boldsymbol{v}^{(j)},\boldsymbol{z}^{(j)},\eta_{\boldsymbol{u}}^{(j)},\eta_{\boldsymbol{v}}^{(j)},\eta_{\boldsymbol{z}}^{(j)}), \tag{6.79}$$

$$\boldsymbol{v}^{(j+1)} = \underset{\boldsymbol{v}}{arg\,min}\, L(\boldsymbol{x}^{(j+1)},\underline{\alpha}^{(j+1)},\boldsymbol{u}^{(j+1)},\boldsymbol{v},\boldsymbol{z}^{(j)},\eta_{\boldsymbol{u}}^{(j)},\eta_{\boldsymbol{v}}^{(j)},\eta_{\boldsymbol{z}}^{(j)}), \tag{6.80}$$

$$\boldsymbol{z}^{(j+1)} = \underset{\boldsymbol{z}}{arg\,min}\, L(\boldsymbol{x}^{(j+1)},\underline{\alpha}^{(j+1)},\boldsymbol{u}^{(j+1)},\boldsymbol{v}^{(j+1)},\boldsymbol{z},\eta_{\boldsymbol{u}}^{(j)},\eta_{\boldsymbol{v}}^{(j)},\eta_{\boldsymbol{z}}^{(j)}). \tag{6.81}$$

Thus, at the jth iteration, instead of (6.73) to (6.76), we execute the following:

$$\boldsymbol{x}^{(j+1)} = \underset{\boldsymbol{x}}{arg\,min}\, \frac{\mu_{\boldsymbol{u}}}{2}\left\|\boldsymbol{u}^{(j)} - \boldsymbol{T}(\underline{\alpha}^{(j)})\boldsymbol{x} - \eta_{\boldsymbol{u}}^{(j)}\right\|_{\Lambda_{\boldsymbol{u}}}^2 + \frac{\mu_{\boldsymbol{z}}}{2}\left\|\boldsymbol{z}^{(j)} - \boldsymbol{x} - \eta_{\boldsymbol{z}}^{(j)}\right\|_{\Lambda_{\boldsymbol{z}}}^2, \tag{6.82}$$

$$\underline{\alpha}^{(j+1)} = \underset{\underline{\alpha}}{arg\,min}\, \mathsf{R}_2(\underline{\alpha}) + \frac{\mu_{\boldsymbol{u}}}{2}\left\|\boldsymbol{u}^{(j)} - \boldsymbol{T}(\underline{\alpha})\boldsymbol{x}^{(j+1)} - \eta_{\boldsymbol{u}}^{(j)}\right\|_{\Lambda_{\boldsymbol{u}}}^2, \tag{6.83}$$

$$\boldsymbol{u}^{(j+1)} = \underset{\boldsymbol{u}}{arg\,min}\, \frac{\mu_{\boldsymbol{u}}}{2}\left\|\boldsymbol{u} - \boldsymbol{T}(\underline{\alpha}^{(j+1)})\boldsymbol{x}^{(j+1)} - \eta_{\boldsymbol{u}}^{(j)}\right\|_{\Lambda_{\boldsymbol{u}}}^2 + \frac{\mu_{\boldsymbol{v}}}{2}\left\|\boldsymbol{v}^{(j)} - \boldsymbol{A}\boldsymbol{u} - \eta_{\boldsymbol{v}}^{(j)}\right\|_{\Lambda_{\boldsymbol{v}}}^2, \tag{6.84}$$

$$\boldsymbol{v}^{(j+1)} = \underset{\boldsymbol{v}}{arg\,min}\, \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{v}\|_{\boldsymbol{W}}^2 + \frac{\mu_{\boldsymbol{v}}}{2}\left\|\boldsymbol{v} - \boldsymbol{A}\boldsymbol{u}^{(j+1)} - \eta_{\boldsymbol{v}}^{(j)}\right\|_{\Lambda_{\boldsymbol{v}}}^2, \tag{6.85}$$

$$\boldsymbol{z}^{(j+1)} = \underset{\boldsymbol{z}}{arg\,min}\, \mathsf{R}(\boldsymbol{z}) + \frac{\mu_{\boldsymbol{z}}}{2}\left\|\boldsymbol{z} - \boldsymbol{x}^{(j+1)} - \eta_{\boldsymbol{z}}^{(j)}\right\|_{\Lambda_{\boldsymbol{z}}}^2, \tag{6.86}$$

$$\eta_{\boldsymbol{u}}^{(j+1)} = \eta_{\boldsymbol{u}}^{(j)} - (\boldsymbol{u}^{(j+1)} - \boldsymbol{T}(\underline{\alpha}^{(j+1)})\boldsymbol{x}^{(j+1)}), \tag{6.87}$$

$$\eta_{\boldsymbol{v}}^{(j+1)} = \eta_{\boldsymbol{v}}^{(j)} - (\boldsymbol{v}^{(j+1)} - \boldsymbol{A}\boldsymbol{u}^{(j+1)}), \tag{6.88}$$

$$\eta_{\boldsymbol{z}}^{(j+1)} = \eta_{\boldsymbol{z}}^{(j)} - (\boldsymbol{z}^{(j+1)} - \boldsymbol{x}^{(j+1)}), \tag{6.89}$$

Out of equations (6.82) to (6.89), only (6.83) is related to the motion estimation. All other equations are related to the motion-compensated image reconstruction.

We first consider the motion estimation step (6.83). This minimization can be separated into N_f image registration problems. Since the minimization is in image domain, it will be much faster than the conventional joint registration, which is in measurement domain. Also, due to the independence between the motion parameters in each frame, we can parallelize the minimization process.

Sub-problems (6.82), (6.84), and (6.85) are same as (6.35), (6.36), and (6.37) from previous section, respectively. Therefore, (6.82) and (6.84) can be solved by the iterative CG solver with previously defined preconditioners, and (6.85) has an analytical solution (6.48).

We can observe that (6.86) is a kind of image denoising problem, which can be easily solved with iterative algorithms like gradient descent. Finally, we present our algorithm in Table 6.8 for solving (6.71).

The parameters are selected by using the same principles from previous section.

6.3.3 Sinogram-based motion estimation for coronary artery

6.3.3.1 Motivation

Since the accuracy of the estimated motion significantly affects the reconstructed images, obtaining precise motion estimates is very important. However, estimating the motion of the entire heart or even just the coronary arteries is a very challenging problem. Many approaches have been proposed to address the problem [63, 106, 122]. However, most of these methods require

Table 6.8: Splitting approach for joint motion estimation and image reconstruction in X-ray CT imaging

1. Select $\boldsymbol{x}^{(0)}$, $\mu_{\boldsymbol{u}}$, $\mu_{\boldsymbol{v}}$, $\mu_{\boldsymbol{z}} > 0$ and set j = 0
2. Set $\boldsymbol{u}^{(0)} = \boldsymbol{T}(\underline{\alpha})\boldsymbol{x}^{(0)}$, $\boldsymbol{v}^{(0)} = \boldsymbol{A}\boldsymbol{u}^{(0)}$, $\boldsymbol{z}^{(0)} = \boldsymbol{x}^{(0)}$, and $\eta_{\boldsymbol{u}}^{(0)} = \eta_{\boldsymbol{v}}^{(0)} = \eta_{\boldsymbol{z}}^{(0)} = 0$

Repeat:

3. Obtain $\boldsymbol{x}^{(j+1)}$ using PCG to (6.82)
4. Obtain $\underline{\alpha}^{(j+1)}$ using iterative algorithm to (6.83)
5. Obtain $\boldsymbol{u}^{(j+1)}$ applying PCG to (6.84)
6. Obtain $\boldsymbol{v}^{(j+1)}$ using (6.48)
7. Obtain $\boldsymbol{z}^{(j+1)}$ applying iterative algorithm to (6.86)
8. $\eta_{\boldsymbol{u}}{}^{(j+1)} = \eta_{\boldsymbol{u}}{}^{(j)} - (\boldsymbol{u}^{(j+1)} - \boldsymbol{T}(\underline{\alpha})\boldsymbol{x}^{(j+1)})$
9. $\eta_{\boldsymbol{v}}{}^{(j+1)} = \eta_{\boldsymbol{v}}{}^{(j)} - (\boldsymbol{v}^{(j+1)} - \boldsymbol{A}\boldsymbol{u}^{(j+1)})$
10. $\eta_{\boldsymbol{z}}{}^{(j+1)} = \eta_{\boldsymbol{z}}{}^{(j)} - (\boldsymbol{z}^{(j+1)} - \boldsymbol{x}^{(j+1)})$
11. Set j = j + 1

Until stopping criterion is satisfied

extra measurements in addition to the short-scan or depend on the initial gated reconstructions. Joint estimation of both the motion parameters and the reconstructed image may compensate for inaccurate initial motion estimates, but having decent initial estimates is still very desirable. We propose a new approach to estimate the motion of the coronary arteries by using only the short scan measurements. The proposed method tracks the motion of the coronary arteries in the sinogram domain, and provides reasonable estimates of their motion. The acquired motion information can be used to obtain reconstructed images with reduced artifacts. However, the main purpose of this method is not perfectly estimating the motion of the coronary arteries, but rather providing good initial estimates for subsequent joint estimation of both the motion parameters and the image. The method was investigated with a simulation of cardiac CT with the XCAT phantom. Simulation results illustrate that our proposed method can provide reasonably good motion estimates that may be useful for initializing the joint estimation of both motion parameters and reconstructed images.

6.3.3.2 Coronary artery motion estimation

Consider the penalized weighted least squares (PWLS) formulation of motion-compensated CT image reconstruction in (6.22) and estimation of motion parameters and image reconstruction formulated as (6.69). Our proposed can be used to provide initial motion estimates for both

frameworks.

The basic idea is to track the motion of each coronary artery using its trajectory in the sinogram. Given a stationary point in the image domain, we know its exact path in the sinogram domain. In turn, from the path of a point in the sinogram, we can estimate its location during the scan. This information is then used to obtain the warp between frames.

We use 3rd generation CT geometry for the derivation of the method, and only consider the 2D fan beam case for simplicity. However, the proposed method can be applied to any other scanner geometry with proper modifications. We also assume that the displacement of the coronary artery is small, and can be modeled as a linear function of the view angle. Finally, we need approximate locations of the center of each coronary artery. Coronary artery centerline extraction methods or manual inputs can be used to obtain this information, which is only used to determine the approximate region-of-interest (ROI) around each coronary artery.

Step 1 To extract the trajectory of each coronary artery from the measurements, it is helpful to remove the projections corresponding to the objects outside the ROI. By reconstructing only the ROI and subtracting it from the fully reconstructed image, we can obtain the objects outside the ROI. We use filtered back-projection (FBP) method to reconstruct these initial images. For coronary arteries near the boundary of the heart, we fill the very low intensity regions such as air with approximate myocardium CT value before the subtraction. This procedure helps the tracking process to be more accurate. Subtracting the re-projection of this non-ROI image from the original measurements will approximately give the projections of only the ROI region (See Fig. 6.16 (a)). Morphological top-hat filtering can be used to remove unnecessary structures to help isolate the coronary artery [52].

Step 2 Given the projections of only the ROI region, we can identify the trajectory of the target coronary artery by using various signal processing techniques. We used the following procedure:

- Apply low-pass or matched filter to the projections.
- Track the maximum value at each view angle.
- Fit the nth order polynomial to obtain trajectory. We use $n = 5$.

With above procedures, we have the trajectory of the center of a coronary artery, $s_\theta(x, y)$ where θ is the view angle (See Fig. 6.16 (b)).

Step 3 From the geometry of the scanner, we have the following relationship:

$$\begin{aligned} s_\theta(x,y) &= Dsd\ \arctan\left(\frac{x_\theta}{Dso-y_\theta}\right)+\epsilon, && (6.90)\\ x_\theta &= x\cos\theta + y\sin\theta, && (6.91)\\ y_\theta &= -x\sin\theta + y\cos\theta, && (6.92) \end{aligned}$$

where x and y are the true locations of the coronary arteries at view, Dsd is the distance from the source to the detector, Dso is the distance from the source to the iso-center, and ϵ is the noise in the measurement. We obtain the following approximation from (6.90):

$$\begin{aligned} z_\theta \triangleq \tan\left(\frac{s_\theta}{Dsd}\right) &\approx \frac{x_\theta}{Dso-y_\theta},\\ (Dso-y_\theta)\,z_\theta &\approx x_\theta,\\ Dsoz_\theta &\approx x_\theta + z_\theta y_\theta. \end{aligned}$$

Let $x = x_0 + tdx$ and $y = y_0 + tdy$ where $t \in [0\ 1]$, x_0 and y_0 are the initial locations of the coronary artery, and dx and dy are the displacement to the final locations:

$$\begin{aligned} Dsoz_\theta &\approx (x_0+tdx)\cos\theta + (y_0+tdy)\sin\theta\\ &+ z_\theta\{-(x_0+tdx)\sin\theta + (y_0+tdy)\cos\theta\}. \end{aligned}$$

We assumed that the coronary arteries moves at a constant speed, i.e. t increases linearly with view angle θ. Our objective function to optimize can be written as follows:

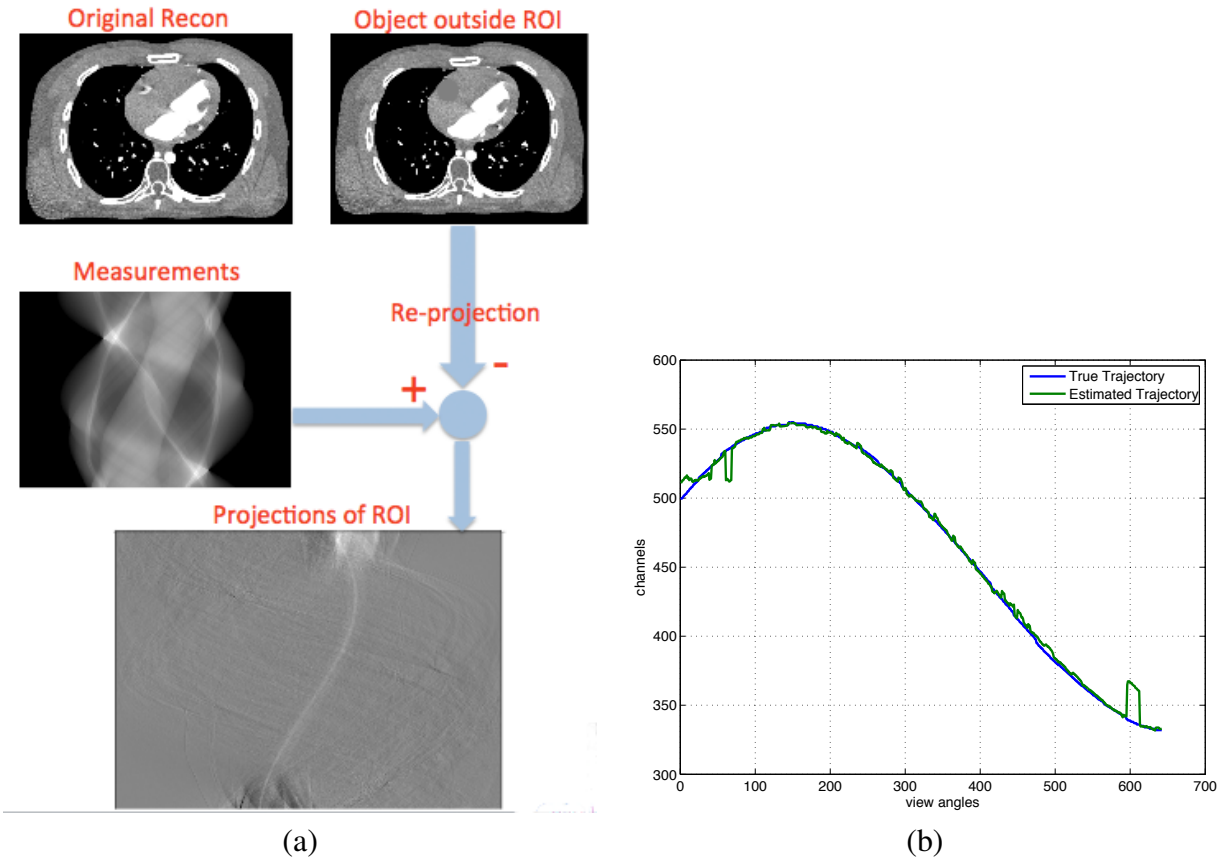
$$\underset{\boldsymbol{u}}{arg\,min}\,\|\boldsymbol{z}-\boldsymbol{B}\boldsymbol{u}\|_1\,, \tag{6.93}$$

where

$$\begin{aligned} \boldsymbol{z} &= [\cdots Dsoz_\theta \cdots]',\\ \boldsymbol{B} &= \begin{bmatrix} & \cos\theta - z_\theta\sin\theta & \\ \cdots & \sin\theta + z_\theta\cos\theta & \cdots \\ & t\cos\theta - tz_\theta\sin\theta & \\ & t\sin\theta + tz_\theta\cos\theta & \end{bmatrix}',\\ \boldsymbol{u} &= \begin{bmatrix} x_0 & y_0 & dx & dy \end{bmatrix}'. \end{aligned}$$

By estimating $\boldsymbol{u}$, we determine the initial location of the coronary artery and its displacement during the scan. Since $\boldsymbol{z}_\theta$ can be noisy, we use L1 regression instead of least squares fitting.

Using the above estimation, we can define the warp between the frames by using the deformation model with thin plate splines (TPS) or B-splines. We used cubic B-splines to model the warp.



(a) (b)

Figure 6.16: (a) Flow chart for step 1 and (b) trajectory of the center of the right coronary artery in the sinogram domain.

6.3.4 Results

6.3.4.1 Sinogram-based motion estimation

The proposed algorithm (6.93) was demonstrated on a 2D CT image reconstruction problem with the XCAT phantom, which provides realistic cardiac motion. Simulated heart rate was 75 bpm, and the case of imperfect cardiac gating was assumed. A short scan of a 3rd-generation fan-beam CT system was simulated, and the system has 888 channels per view spaced 1.0239 mm apart, and 642 evenly spaced view angles. Reconstructed images use pixels $(0.9766mm)^2$ in size with 512^2 voxels.

The approximate location of each coronary artery was manually selected, and a circular region with 15 voxel radius centered at the location was selected as ROI. We assumed seven frames,

$N_f = 7$, and the warp between each frame and the reference frame was obtained using nonrigid image registration and the estimated motion of each coronary artery.

For the regularizer, we used a hyperbola potential function to provide edge-preservation and a certainty-based penalty to obtain more uniform resolution. Poisson noise was added to the measurement, and the weights in the data-fit term in (6.22) were chosen as $w_i = \exp(-[\boldsymbol{A}\boldsymbol{x}]_i)$. The regularization parameter β was selected (without considering the warps) such that the target PSF has a full-width at half-maximum (FWHM) of approximately 1.3 mm.

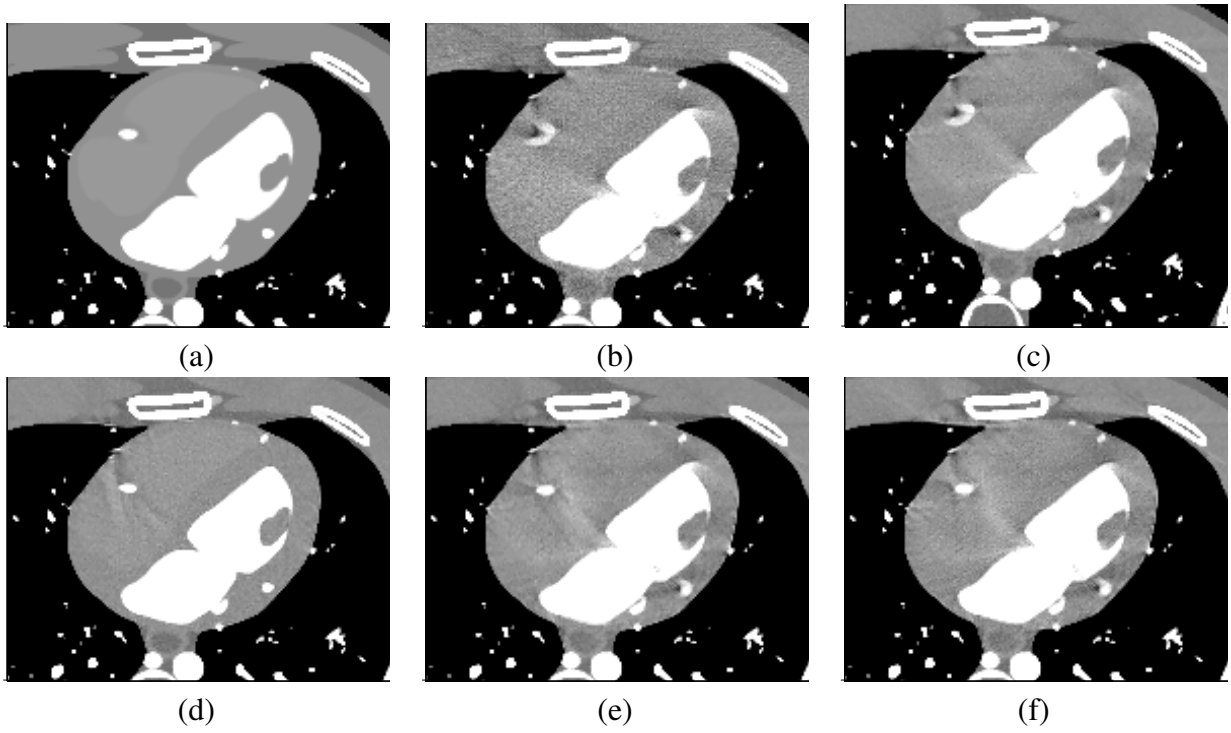


Figure 6.17: Images in the ROI of (a) XCAT phantom; (b) FBP reconstruction; (c) Iterative reconstruction without motion compensation using conjugate gradient (CG); (d) MCIR with the motion estimated directly from the XCAT phantom; (e) MCIR with the motion estimates obtained by the proposed method; (f) Joint estimation results with (e) as a initial starting point.

Fig. 6.17 shows the reconstructed images with various methods. Using FBP or iterative reconstruction without any motion compensation leads to the reconstructed images corrupted with motion artifacts. When accurate motion information is avaiable, MCIR methods can reconstruct the images without motion artifacts, as shown in Fig. 6.17 (d). In this case, the motion estimates were obtained by using nonrigid image registration directly to the phantom.

Our proposed method [Fig. 6.17 (e)] greatly reduces the motion artifacts compared to the conventional reconstruction methods without any compensation. The shape of each coronary artery now became more realistic and close to the truth. However, the reconstructed image still has resid-

	Oracle	FBP	Iterative w/o correction	MCIR	Joint MCIR
1	14.5	66.9	64.9	40.9	41.0
2	28.2	66.9	70.3	35.9	34.2
3	43.1	43.0	40.0	45.0	41.7
4	55.8	39.8	37.5	59.6	61.2

Table 6.9: RMSD (in HU) of different methods around each coronary artery with respect to the true image (Fig. 6.17 (a)). See Fig. 6.17 (a) for the index of each coronary artery.

ual motion artifacts due to suboptimal motion estimation. Fig. 6.17 (f) shows a result from joint estimation of both motion parameters and image, which has further reduced motion artifacts. To quantitatively verify the improvements, we computed the root-mean-squared difference (RMSD) around each coronary artery with respect to the XCAT phantom (See Table 6.9). The results confirm that our proposed methods effectively reduce severe artifacts. However, both our methods and the "oracle" reconstruction [Fig. 6.17 (d)] did not work as desired for the region with little motion. Since our motion model is very simple and tracks only the center of each coronary artery, it may not be enough to precisely capture the local deformation. We'll work further to solve this issue.

6.3.5 Discussion

We presented a new coronary artery motion estimation method for the motion-compensated image reconstruction problem. The proposed method greatly reduced the motion artifacts without requiring any other information or scan padding other than short scan measurements. The estimated motion is limited to the coronary artery, and the estimated motion still suffers from residual motion artifacts because of the suboptimal motion. However, as mentioned in the introduction, the purpose of this method is to provide a good starting point for the following joint estimation step, which showed promise. The proposed method was applied to a 2D case with a digital phantom. Our future work will focus on extending the method to 3-D CT and applying to real patient data. Joint estimation will be more closely investigated to obtain better results.

CHAPTER VII

Conclusions and future work

7.1 Summary

This dissertation addressed various issues of statistical image reconstruction for cardiac CT imaging. The main focus was to improve undesirable properties in the reconstructed image including image artifacts, anisotropic and nonuniform spatial resolution and noise, and poor temporal resolution. We proposed different methods to tackle each of these issues. We used both simulations and clinical data from different scan geometries to evaluate our methods.

Chapter III investigated regularization design methods in the Fourier domain that adjust the regularization parameters, or specifically the directional weighting coefficients, to obtain isotropic and uniform spatial resolution or noise characteristics.

Chapter IV proposed a new concept called "hypothetical geometry" to design regularizers for uniform spatial resolution or noise. The proposed method addresses one main concern of the Fourier based method, which is the local shift-invariance assumption that fails for insufficiently sampled locations.

Chapter V discussed methods to reduce short-scan artifacts in the reconstructed image by exploiting all available information, such as additional measurements or initial reconstructions. The goal of these methods was to reduce short-scan artifacts without degrading temporal resolution within the ROI. Statistical weighting modification method based on the amount of effect of each detector element to insufficiently sampled region was proposed and a metric to evaluate its performance was also introduced. Addition of another regularization term that exploits information from a prior image was investigated. Combination of these methods showed further improvements.

Chapter VI proposed new algorithms to accelerate the statistical image reconstruction for both cases with and without the motion compensation. A new idea called "double surrogates" was introduced to accelerate the family of ordered subsets algorithms, which includes conventional OS, IOT, and variations of OS such as OS + momentum. Variable splitting approach was investigated for motion-compensated image reconstruction problem. We also suggested joint estimation of both

motion parameters and images, and proposed a sinogram-based coronary artery motion estimation method that can provide good initial motion estimates.

7.2 Future work

In this thesis, we proposed various methods to address different aspects of the statistical image reconstruction. Here we introduce possible future works.

We proposed several regularization design methods. Even though these methods showed promise for improving certain image properties like resolution uniformity, they also exhibited limitations. Improving such limitations are possible future works.

For quadratic regularization design methods in Chapter III, we need a systematic method to specify the target response. So far we obtained the target from full-scan geometry and uniform weighting. Such target has isotropic nature, but may not be ideal. Selection of a target requires a separate design process, and is not trivial. Even if we have a process to design a target with desirable properties, the proposed methods cannot precisely match the local response to the target, especially for insufficiently sampled area. Due to limited degree of freedom of the directional weighting design, the variation that can be induced by the regularizer is somewhat limited. Compensating the effect of statistical weightings is possible and demonstrated in previous works. However, the effect of sampling would be hard to compensate. Future works should address this issue. Combining the proposed directional weights design methods with other methods is a possibility. For example, we can apply the hypothetical geometry idea in Chapter IV first, and then apply directional weighting design methods in Chapter III to that extended geometry.

In Appendix B, we briefly discussed a heuristic design to achieve somewhat balanced spatial resolution and noise properties. As mentioned in Chapter 1.2.2, controlling the balance between spatial resolution and noise in user-defined fashion is a desirable ability for a medical imaging modality, especially since each physician may have different preference for diagnostic image quality. All regularization design methods introduced in the thesis may be incorporated into one general design principle. We investigated methods designing different components of the regularizer, and these methods might benefit from each other.

We proposed the hypothetical geometry concept for regularization design in Chapter IV. This idea can also be used for other purposes. For example, the hypothetical geometry can be applied to preconditioner or sampling metric design. It would be interesting to see whether using the hypothetical geometry based preconditioner can further accelerate the variable splitting approach. Since there exist multiple hypothetical geometry for each scanning geometry, the performance of methods using the hypothetical geometry may be further improved if we find more appropriate one.

We investigated methods that remove short-scan artifacts by using all available information. For both statistical weighting method and additional prior regularization method, the main issue is to effectively select the parameter that controls the trade-off. For statistical method, there is a trade-off between the short-scan artifact in the undersampled area and the temporal resolution in the ROI area. For additional regularization method, the trade-off is between the short-scan artifact and undesirable properties of the prior image like poor temporal resolution or noise. To systematically determine the parameter, we need metrics to quantify short-scan artifacts and temporal resolution. For a temporal resolution, we proposed a metric in Chapter V. However, it is not trivial to design a metric for short-scan artifacts since they appear in various forms. Developing a metric that can quantify the short-scan artifact is an interesting and important future work. One possible idea is to use the blur detection algorithm presented in [126]. The algorithm was developed to detect out-of-focus or linear motion blurs in natural images, but the idea of using edge detection and its classification can also be applied to CT images. Furthermore, we can apply the algorithm to small patches to identify the region with artifacts and possibly also their severity. Similar ideas will be investigated.

Regularizers designed in this thesis had the form (2.17) and penalized only the adjacent neighbors of a given voxel to enforce smoothness. However, non-local regularizers may be a good candidate to further improve the reconstructed image quality. Since well-sampled region in the reconstructed image has a good image quality, such information could be used to help improving the image quality of the undersampled region. Investigating non-local regularizers and their design methods is a very interesting future work.

Acceleration methods presented in this thesis can be further improved. Applying double surrogate idea to recently developed ordered-subsets variations could be a future study. In Chapter VI, we did apply double surrogate idea to "OS + momentum" method. However, it was a naive extension, and requires further investigation.

Joint MCIR for cardiac CT is a very challenging problem. Various aspects of the method needs to be addressed for its practical use. Computation cost and initial condition are the most challenging issues among them. We first need to refine the sonogram-based coronary artery motion estimation and extend it to 3D. Since estimating the entire heart motion is very challenging, we want to use coronary artery motion estimates as a starting point. To accelerate joint MCIR, variable splitting approach introduced in Chapter VI needs to be investigated with phantom and clinical experiments. Since variable splitting methods always suffer from parameter selections, investigating ways to select these parameters more rigorously is another future study.

APPENDICES

APPENDIX A

Images for the pre-tuned spatial strength

This appendix compares the pre-tuned spatial strengths introduced in Chapter IV, for the clinical data.

Fig. A.1 shows the pre-tuned spatial strength $\{\kappa_j\}$ of A-REG, defined as (4.1), at xy, xz, and yz planes through the image center. The image shows the entire trans-axial field-of-view (FOV), which is 70 cm. Notice that a cylindrical mask was used.

Fig. A.2 compares the pre-tuned spatial strengths for A-REG, R-REG, and N-REG at both a center slice and an end slice. The pre-tuned spatial strengths contain object information and is smooth. We can verify that the approximation $\kappa_j \approx \kappa_l$ for l within the neighborhood of j is reasonable. Since the measurements are obtained from the short-scan protocol, end slices suffer from insufficient sampling at certain locations. The regularization strengths at such locations becomes overly smoothed and loses object information reflecting their sampling conditions.

Fig. A.4 compares x profiles through the center of the pre-tuned spatial strengths for A-REG, R-REG, and N-REG. We can see that R-REG assigns smaller regularization strength to undersampled region compared to A-REG. N-REG has smaller dynamic range for the regularization strength compared to others due to the square root (4.45).

The approximation in (4.28) was evaluated empirically in Fig. A.3 and Fig. A.5. We can observe that the approximation holds reasonably well. Notice that ripple-like structures in Fig. A.3 (a) are disappeared in Fig. A.3 (b).

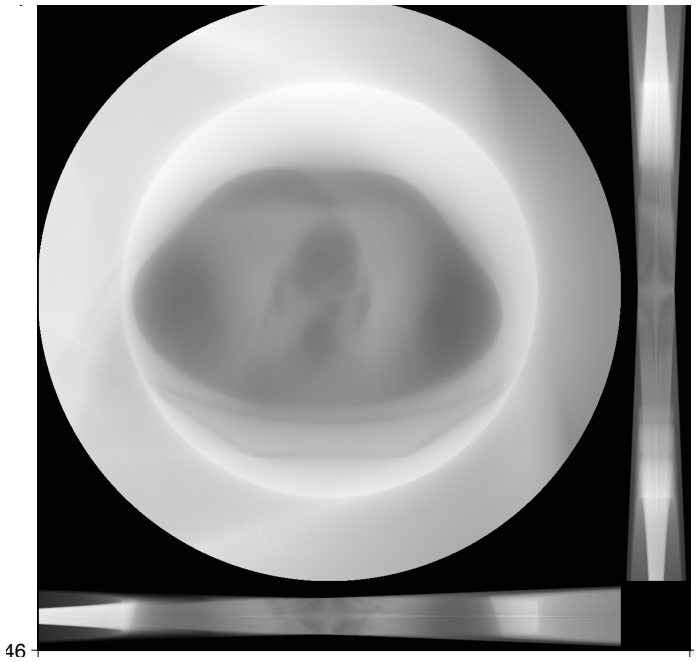
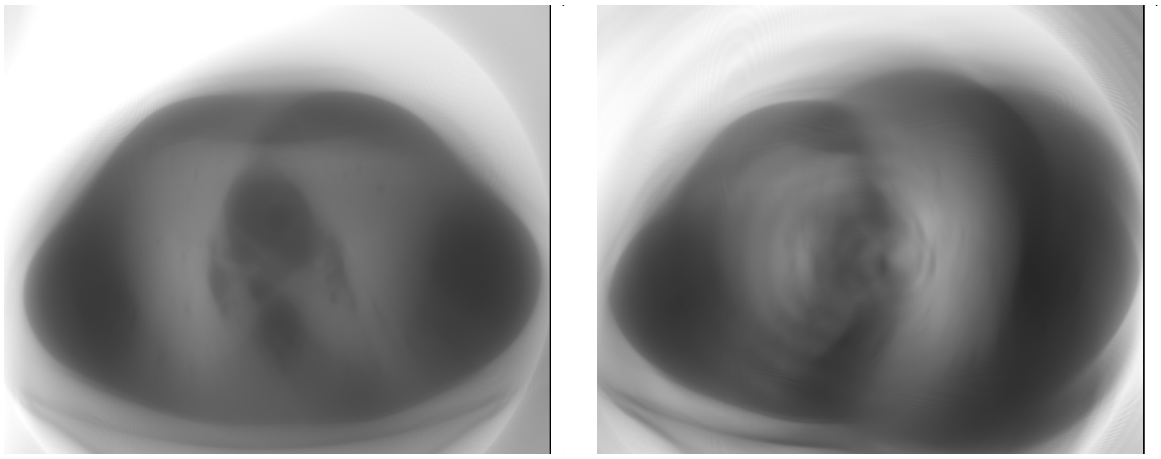
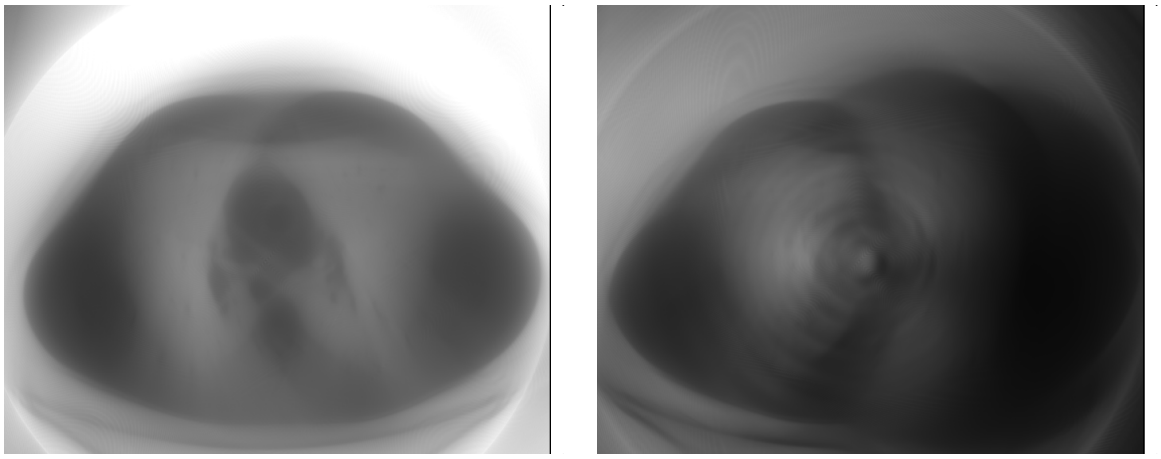


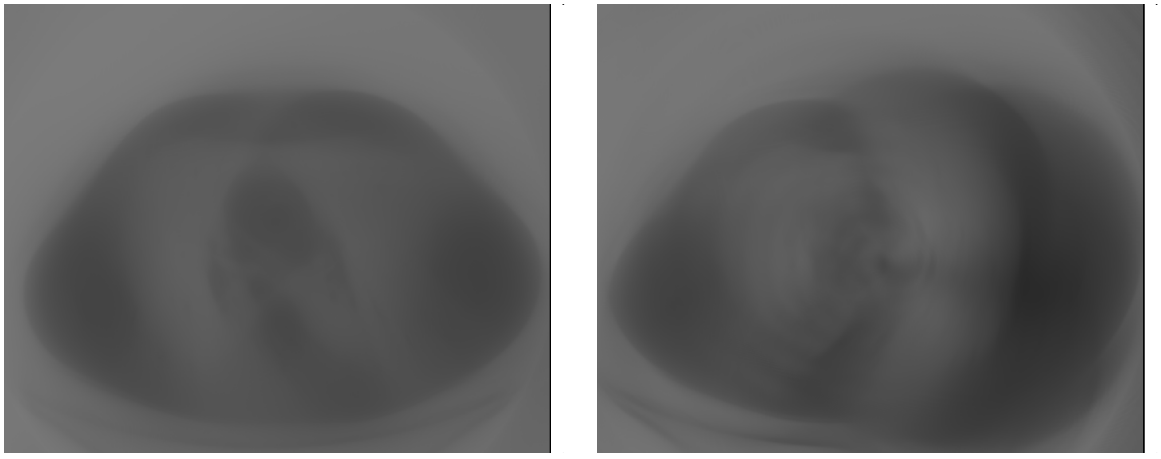
Figure A.1: Middle 3 planes of the pre-tuned spatial strength (4.1) (xy, xz, and yz planes through the image center).



(a) A-REG (4.1)

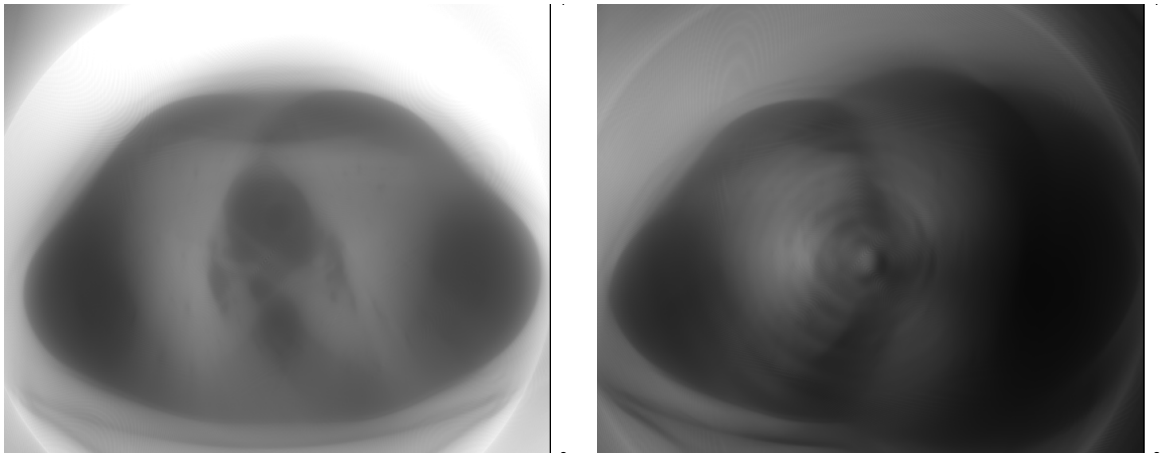


(b) R-REG (4.27)

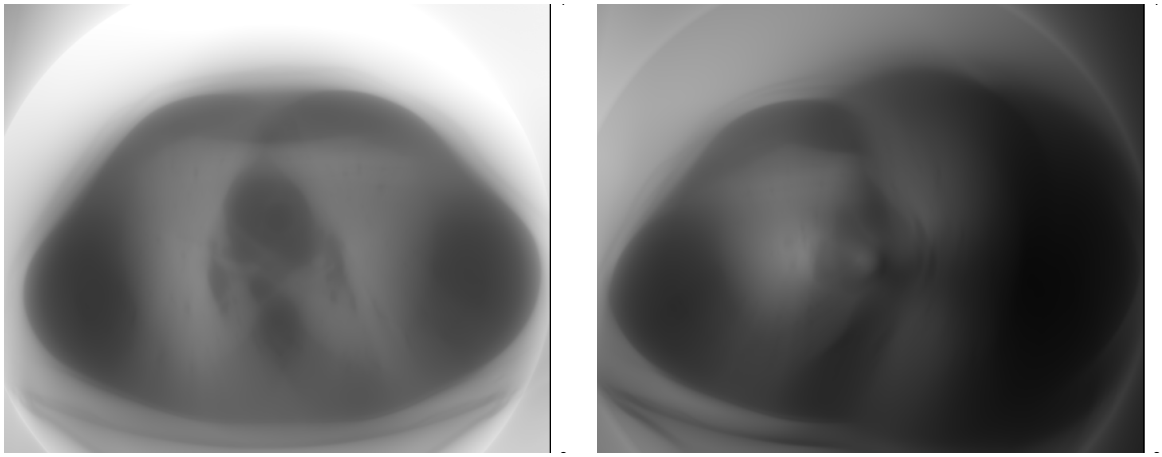


(c) N-REG (4.45)

Figure A.2: Comparison of the pre-tuned spatial strengths for A-REG, R-REG, and N-REG at a center slice (left column) and an end slice (right column). Display range is [0 250].



(a) R-REG (4.27)



(b) Approximated R-REG (4.28)

Figure A.3: Comparison of the pre-tuned spatial strengths for R-REG and its approximation given in (4.27) at a center slice (left column) and an end slice (right column). Display range is [0 250].

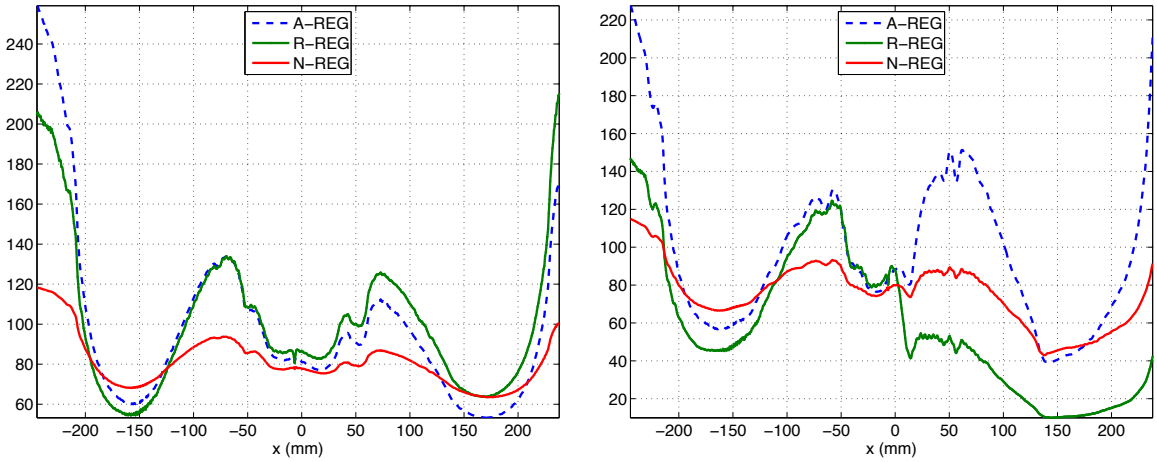


Figure A.4: Comparison of x profiles through the center of the pre-tuned spatial strengths for A-REG, R-REG, and N-REG. From center slice (left) and end slice (right).

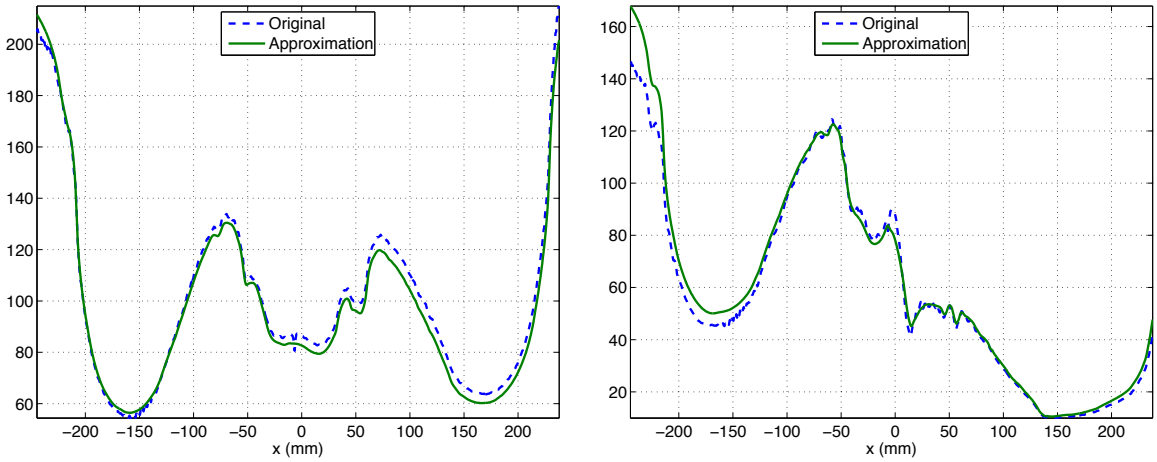


Figure A.5: Comparison of x profiles through the center of the pre-tuned spatial strength (4.27) and its approximation (4.28). From center slice (left) and end slice (right).

APPENDIX B

Supplementary material for regularization designs using the hypothetical geometry

This appendix extends the regularization designs investigated in Chapter IV, and presents the simulation and clinical experiments that supports the results in Chapter IV or were excluded in Chapter IV due to their repetitive nature.

B.1 Selection of the hypothetical geometry

In Chapter IV, new regularizers for uniform resolution or noise were derived using a hypothetical geometry $\boldsymbol{G}$, which was assumed to be an ideal and fully sampled geometry. In practice, the choice of $\boldsymbol{G}$ affects how closely the regularizer can achieve the uniformity of resolution or noise. For the experiments in Chapter IV, $\boldsymbol{G}$ was chosen to be a hypothetical cone-beam CT geometry with large enough number of detector rows and 360° orbit, and the voxel dependent factor s_j was assumed to be 1 for every location. Because of the cone-angle, such hypothetical geometry $\boldsymbol{G}$ does not provide perfectly shift-invariant $\boldsymbol{G}'\boldsymbol{G}$, even though it is "more shift-invariant" compared to $\boldsymbol{A}'\boldsymbol{A}$. This geometry with extended views and rows was one of the most intuitive choices for $\boldsymbol{G}$, and now we propose an alternative another decomposition for $\boldsymbol{G}$.

We let the hypothetical geometry described above as $\boldsymbol{A}_{\text{ext}} = \{\hat{a}_{ij}\}$, then it can be further factorized as follows

$$\hat{a}_{ij} = \bar{a}_{ij}\hat{s}_j, \tag{B.1}$$

so that we obtain

$$\bar{a}_{ij} = \frac{\hat{a}_{ij}}{\hat{s}_j}, \tag{B.2}$$

Table B.1: Acronym suffixes for regularizers

Suffix	Description
-1	regularizers designed by selecting the hypothetical geometry with extended views and rows as $\boldsymbol{G}$ and $s_j = 1, \forall j$
-2	regularizers designed by using the system matrix factorization in (B.4)

where $\hat{s}_j$ is a voxel-dependent factor to make $\bar{\boldsymbol{A}} = \{\bar{a}_{ij}\}$ a better choice for $\boldsymbol{G}$. Intuitively, we want to design $\{\hat{s}_j\}$ such that $\bar{\boldsymbol{A}}'\bar{\boldsymbol{A}}$ has constant diagonal elements. Defining

$$\hat{s}_j = \sqrt{\sum_{i=1}^{N_\mathrm{g}} \hat{a}_{ij}^2} \tag{B.3}$$

yields $[\bar{\boldsymbol{A}}'\bar{\boldsymbol{A}}]_{jj} = 1, \ \forall j$. Using above designs, we obtain the following matrix representation for the system matrix $\boldsymbol{A}$

$$\boldsymbol{A} = \boldsymbol{P}\boldsymbol{G}\boldsymbol{D}[s_j], \tag{B.4}$$

where $\boldsymbol{P}$ is defined in Chapter IV, $g_{ij} = \bar{a}_{ij}$ and $s_j = \hat{s}_j$. The Fisher information matrix can be approximated as (20) with the following new expressions for (4.11) and (4.12):

$$\eta_j = \sqrt{\frac{\sum_{i=1}^{n_\mathrm{d}} g_{ij}^2 w_i}{\sum_{i=1}^{N_\mathrm{g}} g_{ij}^2}} = \sqrt{\sum_{i=1}^{n_\mathrm{d}} g_{ij}^2 w_i}, \tag{B.5}$$

$$\lambda_j = s_j \eta_j = \sqrt{\sum_{i=1}^{n_\mathrm{d}} \hat{a}_{ij}^2 w_i} = \sqrt{\sum_{i=1}^{n_\mathrm{d}} a_{ij}^2 w_i}. \tag{B.6}$$

Notice that (B.6) no longer depends on the hypothetical geometry $\boldsymbol{G}$ and requires less computation compared to both the original certainty (4.1) and the modified certainty (4.12).

New approximation (B.4) for the Fisher information matrix leads to different regularizers from those we used for the experiments in Chapter IV (see Table 4.1 and Table B.1 for acronyms). The main purpose of this supplementary material is to compare these regularizers obtained from different system matrix factorization. Since the derivation for both proposed regularizers in Chapter IV is general, new regularizers are readily obtained from (4.25) and (4.44).

B.1.1 Regularizer for uniform spatial resolution

New regularizer for uniform spatial resolution is obtained by substituting $\boldsymbol{G}$ from (B.4) and (B.6) into (4.25). Using the shift-invariant approximation (4.26), the new regularizer for spatial

resolution uniformity (hereafter R-REG-2) can be written as (2.17) with

$$\hat{\kappa}_j = \sqrt{\sum_{i=1}^{n_\mathrm{d}} a_{ij}^2 w_i} \approx \gamma \sqrt{\sum_{i=1}^{n_\mathrm{d}} a_{ij} w_i}, \tag{B.7}$$

where the approximation holds for a proper scaling factor γ. The approximation scaling factor γ can be obtained manually or by matching both sides of (B.7) at a reference point, *i.e.*,

$$\gamma = \sqrt{\left[\sum_{i=1}^{n_\mathrm{d}} a_{ij}^2 w_i\right]_{j=j_\mathrm{ref}} / \left[\sum_{i=1}^{n_\mathrm{d}} a_{ij} w_i\right]_{j=j_\mathrm{ref}}}, \tag{B.8}$$

where j_ref indicates the lexicographical index of the reference location.

The new regularizer R-REG-2 (B.7) does not have a normalization or a denominator, which also leads to decreased regularization strength for under-sampled region compared to the original certainty (4.1). Both designs (4.27) and (B.7) are expected to have similar effects to the reconstructed image, but possibly with different amount.

B.1.2 Regularizer for uniform noise property

Substituting (B.4) into (4.6) leads to the following new expressions for (4.35) and (4.36)

$$\bar{\lambda}_j \triangleq \sqrt{\sum_{i=1}^{n_\mathrm{d}} a_{ij}^2 \hat{w}_i v_i}, \quad \bar{\Lambda} \triangleq \boldsymbol{D}[\bar{\lambda}_j], \tag{B.9}$$

$$\hat{\lambda}_j \triangleq \sqrt{\sum_{i=1}^{n_\mathrm{d}} a_{ij}^2 \hat{w}_i v_i^2}, \quad \hat{\Lambda} \triangleq \boldsymbol{D}[\hat{\lambda}_j]. \tag{B.10}$$

Substituting $\boldsymbol{G}$ from (B.4), (B.9) and (B.10) into (4.45) yields a new regularizer for noise uniformity (hereafter N-REG-2).

B.2 Comparison between regularizers from different system matrix factorizations

B.2.1 Physical Phantom Experiment

We used the GE performance phantom to compare the resolution properties of the reconstructed image from different regularizations. The phantom was scanned and reconstructed as described in Chapter IV. The GEPP has a tunsten wire in water as indicated in Fig. B.1, and its peak values in

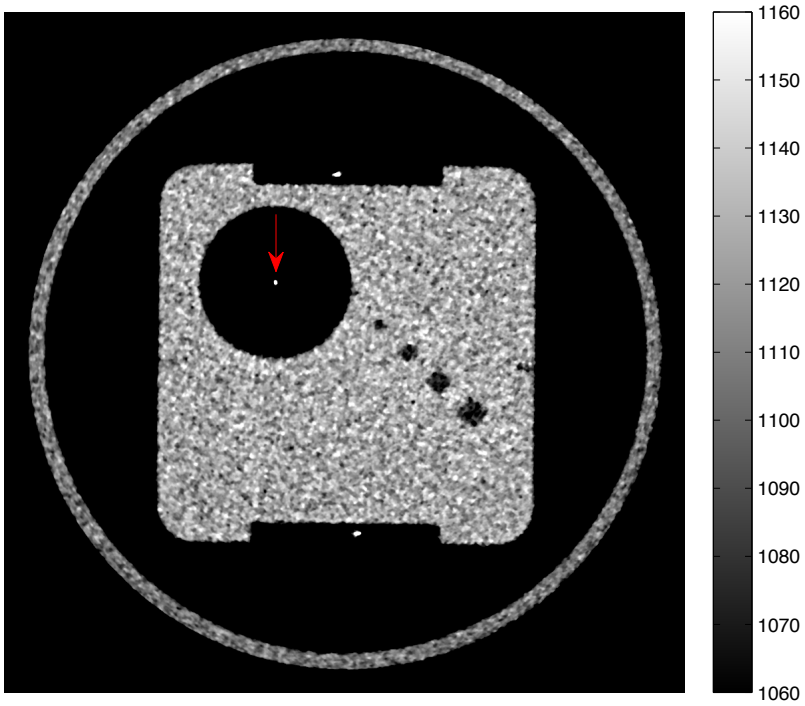


Figure B.1: The GEPP phantom used for quantitative comparison of regularizers. Red arrow indicate tungsten wire used for spatial resolution comparison.

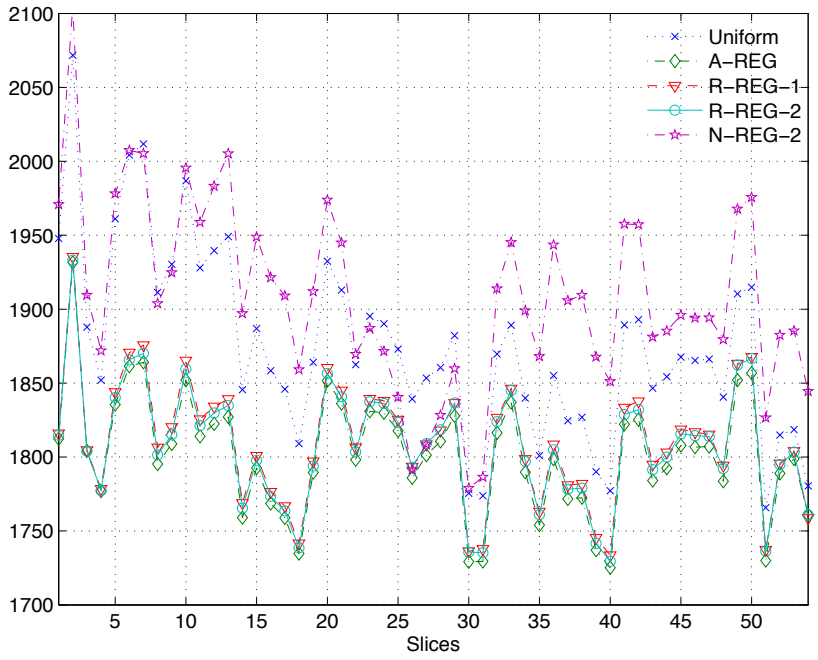


Figure B.2: Comparison of the peak value of the tungsten wire in GEPP for the following regularizers: Uniform, A-REG, R-REG-1, R-REG-2, and N-REG-2.

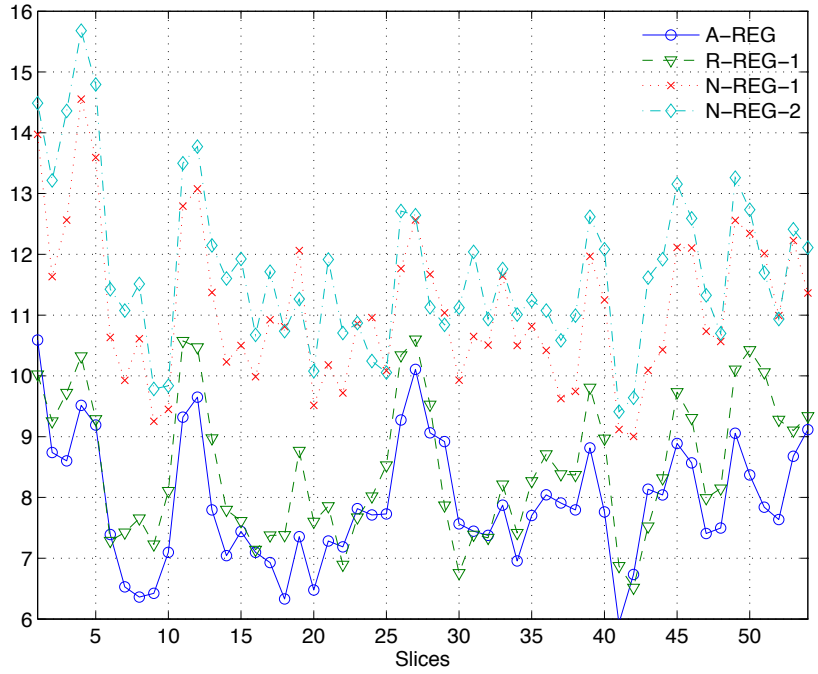


Figure B.3: Comparison of the noise standard deviation in a homogeneous region of the GEPP, which corresponds to the left phantom wall in Table B.4, for the following regularizers: A-REG, R-REG-1, N-REG-1, and N-REG-2.

Table B.2: Average CRC mismatch (4.46) for selected 6 locations across slices (Units: %). See Fig. 4.2 in Chapter IV for the index of impulse locations.

	Locations					
	1	2	3	4	5	6
A-REG (Fig. B.10)	32.6	45.3	24.8	47.2	31.9	25.0
R-REG-1 (Fig. 4.4)	7.5	14.9	3.9	20.0	5.9	7.2
R-REG-2 (Fig. B.10)	11.0	18.7	3.9	23.3	9.7	9.4
A-REG (Fig. B.12)	31.3	44.1	21.9	47.9	33.4	20.6
R-REG-1 (Fig. B.8)	9.5	14.3	2.9	21.5	9.8	7.1
R-REG-2 (Fig. B.12)	11.5	18.1	2.5	24.8	11.9	7.4

Table B.3: Average CRC mismatch (4.46) within each slice, respectively (Units: %).

	Averages			
	Overall	Center Slice	1st slice of ROI	Outside ROI
A-REG (Fig. 4.4)	34.5	(a) 12.3	(b) 33.7	(c) 46.8
R-REG-1 (Fig. 4.4)	9.9	(d) 8.7	(e) 11.6	(f) 7.4
R-REG-2 (Fig. B.10)	12.7	(d) 12.1	(e) 14.4	(f) 10.4
A-REG (Fig. B.12)	33.2	(a) 11.9	(b) 34.5	(c) 44.9
R-REG-1 (Fig. B.8)	10.9	(d) 10.5	(e) 14.6	(f) 9.1
R-REG-2 (Fig. B.12)	12.7	(d) 12.9	(e) 17.0	(f) 9.2

each slice of the reconstructed image will be used to assess resolution uniformity.

In Fig. B.2, we compare the peak value of the tungsten wire profile at every ROI slice for the following regularizers: Uniform, A-REG, R-REG-1,R-REG-2, and N-REG-2. Notice that on center slices all the regularizers generate similar peak values for the tungsten wire due to matched regularization strength at the isocenter. Uniform regularizer and proposed regularizer N-REG-2 for noise uniformity induce non-uniform spatial resolution characteristics to the reconstructed image, which result in widely varying peak values of the tungsten wire. The standard deviation in the peak value is approximately 64 HU for these regularizations. On the other hand, the other 3 regularizers aimed for spatial resolution uniformity show much better performances as shown in Fig. B.2, and have approximately 40 HU standard deviation in the peak value. Due to small FOV (23 cm) and its location, the sampling condition of the tungsten wire is approximately the same for every slices, thus both proposed regularizers R-REG-1 and R-REG-2 perform almost the same as A-REG.

Fig. B.3 compares the noise standard deviation in a homogeneous region of the GEPP, which corresponds to the left phantom wall in Table B.4, for the following regulairzers: A-REG, R-REG-1, N-REG-1 and N-REG-2. A-REG and R-REG-1 show over-regularizing behavior for given location. On the other hand, both proposed regularizers for uniform noise, N-REG-1 and N-REG-2, show improved noise uniformity. Overall performances of N-REG-1 and N-REG-2 are very similar, but show differences for some slices.

B.2.2 XCAT Phantom Simulation

We used the XCAT phantom to illustrate the performance of the proposed regularizer R-REG-2 (B.7). The same simulation settings as in Chapter IV have been used. The improvements in the spatial resolution uniformity were demonstrated by comparing the impulse responses at different voxel locations obtained by using A-REG (4.1) to that of the proposed regularizer R-REG-2.

Results for both quadratic and edge-preserving regularizations are presented from Fig. B.9 to Fig. B.12. We also present the result for R-REG-1 with edge-preserving regularization in Fig. B.7 and Fig. B.8. Similar to the results with R-REG-1 (4.27), the proposed regularizer R-REG-2 (B.7) effectively generate spatially uniform resolution in terms of the constant recovery constant (CRC). Table B.2 and Table B.3 compare CRC mismatch (4.46) of A-REG, R-REG-1, and R-REG-2. R-REG-2 improves the average CRC mismatch by approximately 20% throughout the image volume. Both R-REG-1 and R-REG-2 show comparable performances of improving the resolution uniformity, but R-REG-1 is slightly better.

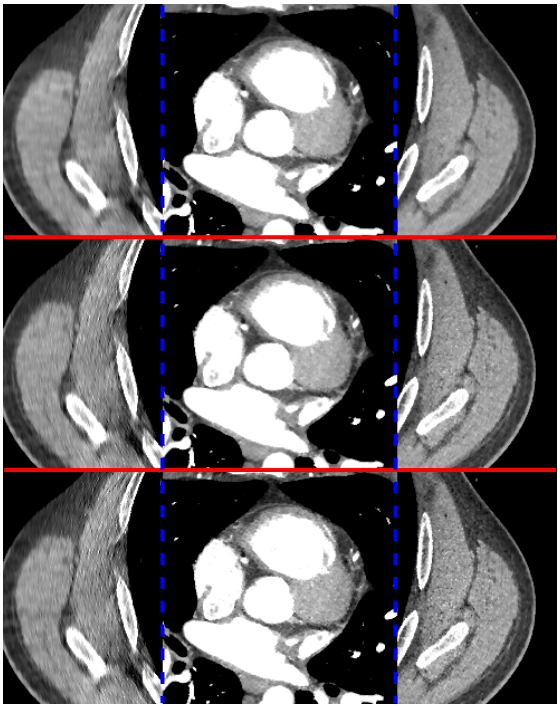


Figure B.4: Comparison of reconstructed images obtained from A-REG (top), R-REG-1 (middle) and R-REG-2 (bottom) at 3 selected locations (separeted by blue dash lines) on the last slice of ROI. Display window is [800 1200] (HU).

B.2.3 Real Clinical Data

Same clinical cardiac CT scan used in Chapter IV was reconstructed with various regularizers. Fig. B.4 compares reconstructed images obtained from both proposed regularizers for resolution uniformity, R-REG-1 and R-REG-2. The reconstructed image using A-REG was presented as a reference. Reconstructed image from R-REG-2 is slightly more shaper and noisier compared to that of R-REG-1 on end slices, but they show very similar image characteristics overall. Fig. B.5 compares reconstructed images from proposed uniform noise regularizers, N-REG-1 and N-REG-2, and they have very comparable visual image quality.

B.3 Compromise Regularizer

As illustrated on previous sections, each proposed regularizer, either aiming for uniform resolution or noise characteristics, reasonably achieves its goal. However, due to the trade-off between these image quality properties, using one of these regularizations may not provide the “most desirable image”, especially near the end slices with sampling issues. Since the readability of the reconstructed image depend on many factors, including both spatial resolution and noise, it is hard to design the optimal regularizer using one criterion. Furthermore, clinicians may have different

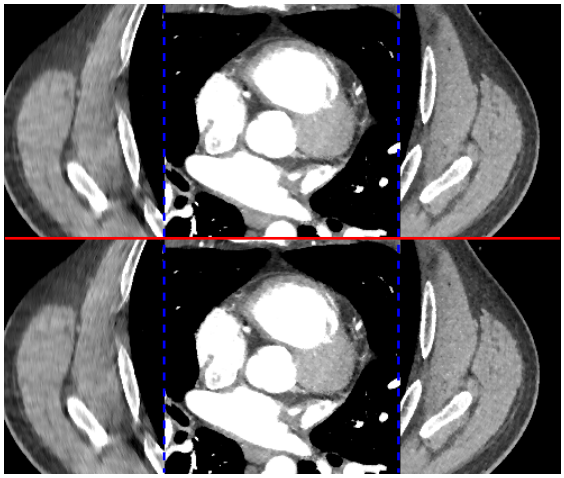


Figure B.5: Comparison of reconstructed images obtained from both uniform noise regularizers, N-REG-1 (top) and N-REG-2 (bottom), at 3 selected locations (separeted by blue dash lines) on the 1st slice of ROI. Display window is [800 1200] (HU).

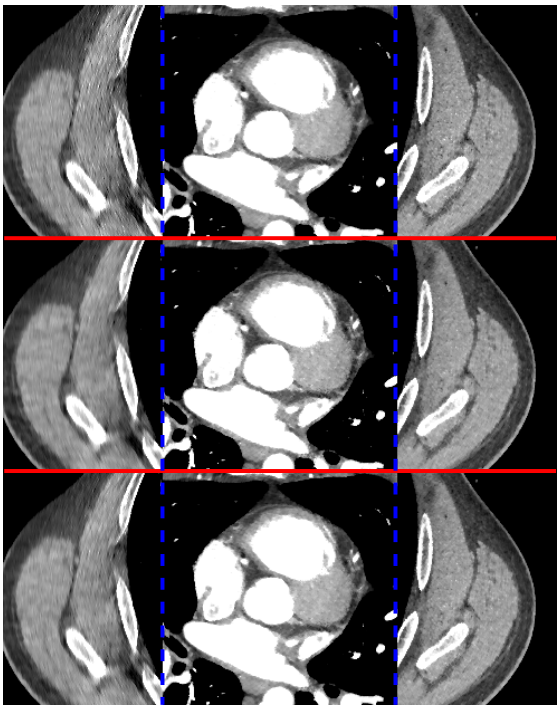


Figure B.6: Comparison of reconstructed images obtained using R-REG-1 (top), N-REG-1 (middle) and compromise regularizer (bottom) at 3 selected locations (separated by blue dash lines) on the last slice of ROI. Display window is [800 1200] (HU).

Table B.4: Comparison of the noise standard deviation for different regularizers method. All values are in Hounsfield units (HU).

	Uniform	A-REG	R-REG-1	R-REG-2	N-REG-1	N-REG-2	Compromise
Plexiglas™ Insert (center)	14.0	13.7	13.7	13.7	13.6	13.3	13.2
Plexiglas™ Insert (right)	18.8	16.0	16.2	15.8	16.4	16.4	15.8
Plexiglas™ Insert (left)	15.1	14.4	14.5	14.7	14.9	14.7	14.2
Phantom Wall (left)	15.4	9.3	9.5	10.0	11.8	12.7	9.2
Phantom Wall (right)	16.9	8.8	6.8	7.4	10.1	10.1	8.3
Phantom Wall (top)	17.4	8.7	8.8	7.6	10.7	10.6	8.6
Phantom Wall (bottom)	16.1	7.7	8.8	9.0	10.9	10.9	9.6
Average	16.2 ± 1.6	11.2 ± 3.4	11.2 ± 3.6	11.2 ± 3.5	12.6 ± 2.4	12.7 ± 2.3	11.3 ± 3.1

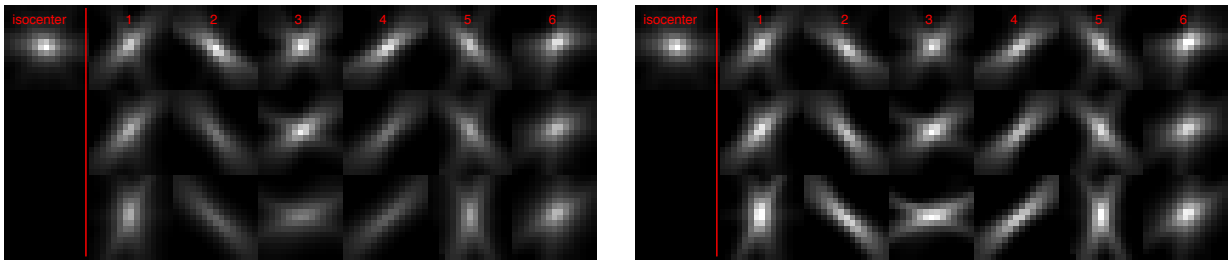
preferences for the appearance of the reconstructed image, complicating the regularization design process.

We investigated a "compromise" regularizer whose certainty function κ_j is the arithmetic average of (4.27) and (4.45). This simple extension of proposed methods gives a sub-optimal solution for the trade-off between spatial resolution and noise property.

Table B.4 shows that the compromise regularizer provides the noise uniformity in between those of R-REG-1 and N-REG-1, as expected. Fig. B.6 compares the reconstructed images from following regularizers: R-REG-1, N-REG-1, and compromise regularizer. When undersampled region (left column in Fig. B.6) is compared, the reconstructed image from compromise reconstruction shows somewhat balanced image quality in spatial resolution and noise properties. However, such improvements are rather subtle.

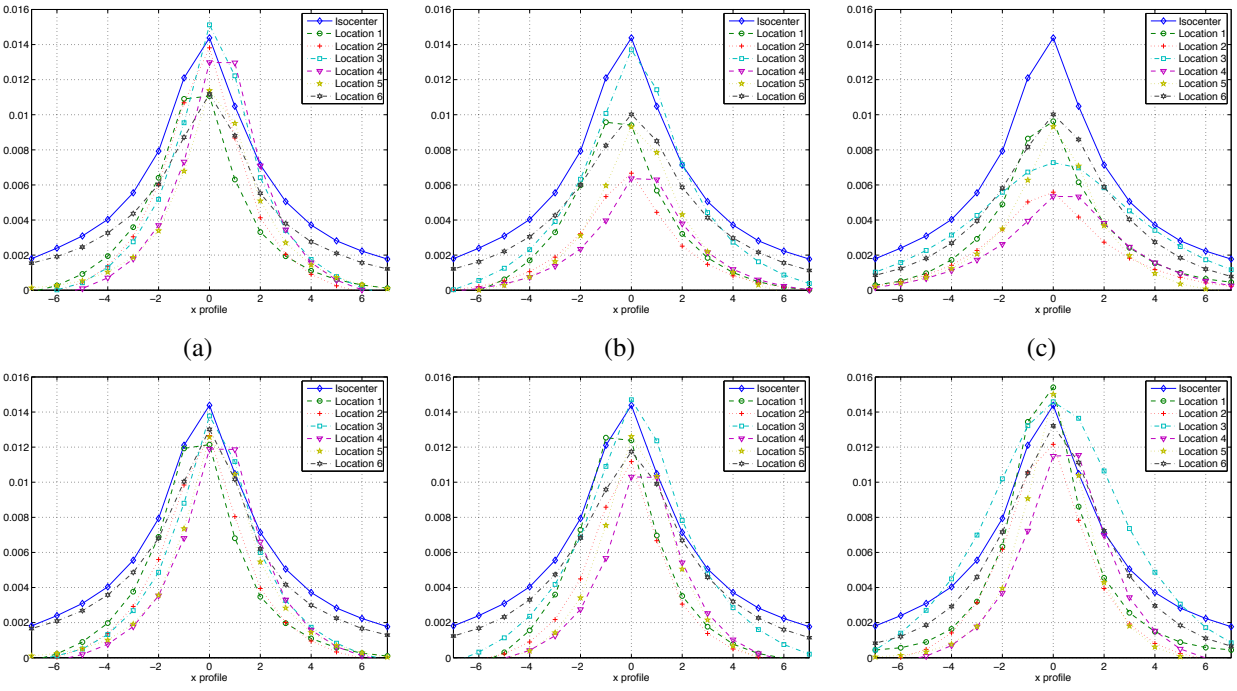
B.4 Summary

The choice of the hypothetical geometry $\boldsymbol{G}$ affects the performance of the regularizers that we proposed in Chapter IV. In addition to an intuitive selection we made in Chapter IV, we presented another factorization for the system model in this supplement, which led to new expressions for the proposed regularizers. These new regularizers, R-REG-2 and N-REG-2, showed comparable per-



(a) A-REG (b) R-REG

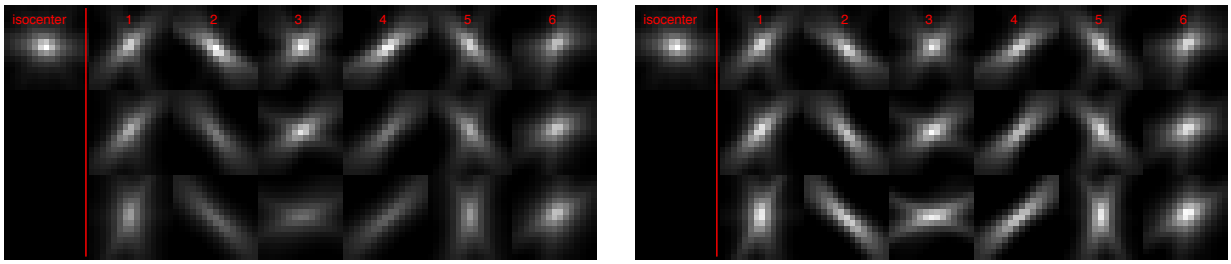
Figure B.7: Comparison of xy plane through the center of each local impulse response at selected location (see Fig. 4.2 for the index of locations). Edge-preserving potential function was used. Top row is from a center slice, middle row is from 1st slice of ROI, and bottom row is from outside ROI. (a) Regularization with original aggregated certainty (4.1) (A-REG) (b) Regularization with proposed certainty (4.27) (R-REG-1).



(a) (b) (c)

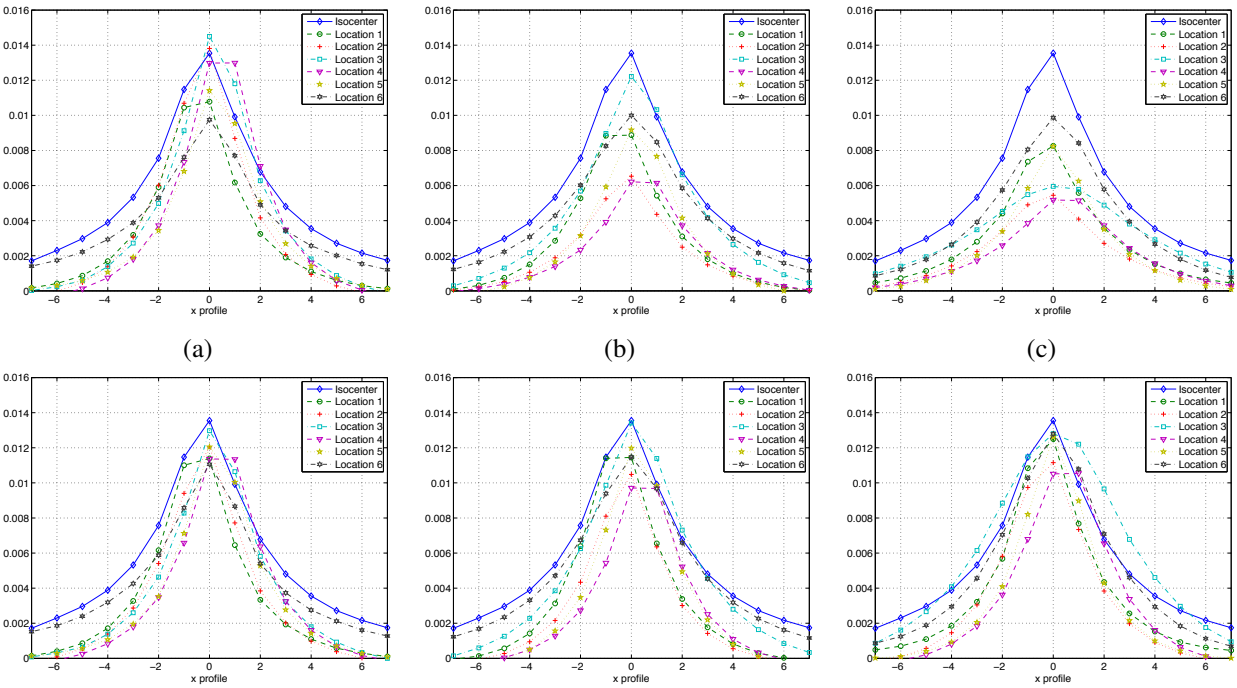
(d) (e) (f)

Figure B.8: Comparison of x profiles through the center of each impulse response in Fig. B.7. Left column is from a center slice, middle column is from 1st slice of ROI, and right column is from outside of ROI. Top and bottom rows represent the regularizers A-REG (4.1) and the proposed R-REG-1 (4.27), respectively. (a) A-REG, center slice (b) A-REG, 1st slice of ROI (c) A-REG, outside ROI (d) R-REG-1, center slice (e) R-REG-1, 1st slice of ROI (f) R-REG-1, outside ROI.



(a) A-REG (b) R-REG-2

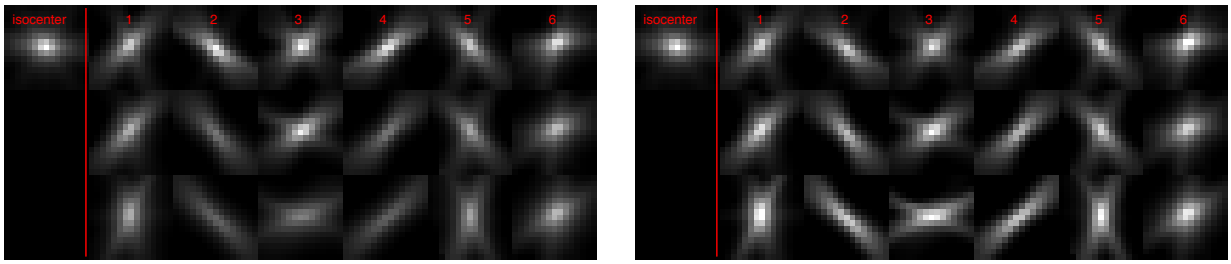
Figure B.9: Comparison of xy plane through the center of each local impulses responses at selected location (see Fig. 4.2 in Chapter IV for the index of locations). Quadratic penalty function was used. Top row is from a center slice (blue line in Fig. 4.2 in Chapter IV), middle row is from 1st slice of ROI (red line), and bottom row is from outside ROI (green line). (a) Regularization with original aggregated certainty (4.1) (A-REG) (b) Regularization with proposed certainty (B.7) (R-REG-2).



(a) (b) (c)

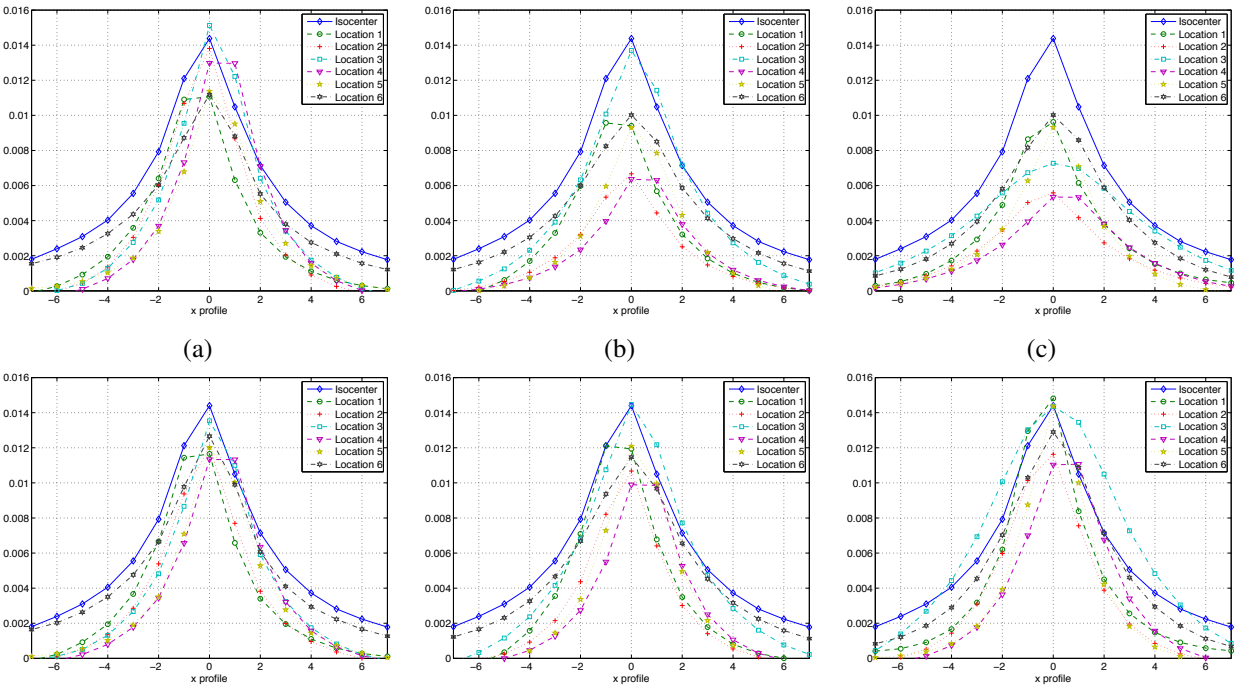
(d) (e) (f)

Figure B.10: Comparison of x profiles through the center of each local impulse response in Fig. B.9. Left column is from a center slice, middle column is from 1st slice of ROI, and right column is from outside of ROI. Top and bottom rows represent the regularizers A-REG (4.1) and the proposed R-REG-2 (B.7), respectively. (a) A-REG, center slice (b) A-REG, 1st slice of ROI (c) A-REG, outside ROI (d) R-REG-2, center slice (e) R-REG-2, 1st slice of ROI (f) R-REG-2, outside ROI.



(a) A-REG (b) R-REG-2

Figure B.11: Comparison of xy plane through the center of each local impulses responses at selected location (see Fig. 4.2 in Chapter IV for the index of locations). Edge-preserving penalty function was used. Top row is from a center slice (blue line in Fig. 4.2 in Chapter IV), middle row is from 1st slice of ROI (red line), and bottom row is from outside ROI (green line). (a) Regularization with original aggregated certainty (4.1)] (A-REG) (b) Regularization with proposed certainty (B.7) (R-REG-2).



(a) (b) (c)

(d) (e) (f)

Figure B.12: Comparison of x profiles through the center of each local impulse response in Fig. B.11. Left column is from a center slice, middle column is from 1st slice of ROI, and right column is from outside of ROI. Top and bottom rows represent the regularizers A-REG (4.1) and the proposed R-REG-2 (B.7), respectively. (a) A-REG, center slice (b) A-REG, 1st slice of ROI (c) A-REG, outside ROI (d) R-REG-2, center slice (e) R-REG-2, 1st slice of ROI (f) R-REG-2, outside ROI.

formances in terms of achieving uniform spatial resolution or noise properties in the reconstructed image compared to those presented in Chapter IV, R-REG-1 and N-REG-1. It is hard to conclude which factorization is a better choise because R-REG-1 and N-REG-2 obtained bettern resolution and noise uniformities compared to R-REG-2 and N-REG-1, respectively.

We also investigated a compromise regularizer as a starting point for investigating the trade-off between spatial resolution and noise. The proposed regularizer obtained somewhat balanced spatial resolution and noise properties. However, such compromise regularizers require more detailed investigation.

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] H. Erdoğan and J. A. Fessler. Ordered subsets algorithms for transmission tomography. *Phys. Med. Biol.*, 44(11):2835–51, November 1999.

[2] M. Kachelrieß, S. Schaller, and W. A. Kalender. Advanced single-slice rebinning in cone-beam spiral CT. *Med. Phys.*, 27(4):754–72, April 2000.

[3] M. V. Afonso, José M Bioucas-Dias, and Mário A T Figueiredo. Fast image recovery using variable splitting and constrained optimization. *IEEE Trans. Im. Proc.*, 19(9):2345–56, September 2010.

[4] M. V. Afonso, J. M. Bioucas-Dias, and M. A. T. Figueiredo. An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems. *IEEE Trans. Im. Proc.*, 20(3):681–695, March 2011.

[5] S. Ahn and J. A. Fessler. Globally convergent image reconstruction for emission tomography using relaxed ordered subsets algorithms. *IEEE Trans. Med. Imag.*, 22(5):613–26, May 2003.

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

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

[8] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.*, 2(1):183–202, 2009.

[9] D. P. Bertsekas. Multiplier methods: A survey. *Automatica*, 12(2):133–45, March 1976.

[10] D. P. Bertsekas. A new class of incremental gradient methods for least squares problems. *SIAM J. Optim.*, 7(4):913–26, November 1997.

[11] R. Bhagalia, J. V. Miller, and A. Roy. Bi-directional labeled point matching. In *Proc. IEEE Intl. Symp. Biomed. Imag.*, pages 380–3, 2010.

[12] D. Blatt, A. O. Hero, and H. Gauchman. A convergent incremental gradient method with a constant step size. *SIAM J. Optim.*, 18(1):29–51, 2007.

[13] M. Blume, A. Martinez-Moller, A. Keil, N. Navab, and M. Rafecas. Joint reconstruction of image and motion in gated positron emission tomography. *IEEE Trans. Med. Imag.*, 29(11):1892–906, November 2010.

[14] F. E. Boas and D. Fleischmann. CT artifacts: causes and reduction techniques. *Imaging in Medicine*, 4(2):229–40, April 2012.

[15] W. Byrne and A. Gunawardana. Comments on "Efficient training algorithms for HMMs using incremental estimation". *IEEE Trans. Speech & Audio Proc.*, 8(6):751–4, November 2000.

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

[17] J. H. Cho and J. A. Fessler. Accelerating ordered-subsets image reconstruction for X-ray CT using double surrogates. In *Proc. SPIE Medical Imaging 2012: Phys. Med. Im.*, page 83131X, 2012.

[18] J. H. Cho and J. A. Fessler. Motion-compensated image reconstruction for cardiac CT with sinogram-based motion estimation. In *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, 2013.

[19] J. H. Cho and J. A. Fessler. Quadratic regularization design for 3D axial CT. In *Proc. Intl. Mtg. on Fully 3D Image Recon. in Rad. and Nuc. Med*, pages 78–81, 2013.

[20] J. H. Cho and J. A. Fessler. Quadratic regularization design for 3D axial CT: Towards isotropic noise. In *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, 2013.

[21] J. H. Cho and J. A. Fessler. Regularization designs for uniform spatial resolution and noise properties in statistical image reconstruction for 3D x-ray CT. *IEEE Trans. Med. Imag.*, 2014. Submitted.

[22] J. H. Cho and J. A. Fessler. Uniform spatial resolution for statistical image reconstruction for 3D cone-beam CT. In *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, 2014. To appear.

[23] J. H. Cho, D. Pal, J-B. Thibault, and J. A. Fessler. Short-scan artifact removal for 3D axial cone-beam CT using all acquired views. In *Proc. SPIE Medical Imaging 2015: Phys. Med. Im.*, 2015. Submitted. MI101-238.

[24] J. H. Cho, S. Ramani, and J. A. Fessler. Alternating minimization approach for multi-frame image reconstruction. In *IEEE Workshop on Statistical Signal Processing*, pages 225–8, 2012.

[25] J. H. Cho, S. Ramani, and J. A. Fessler. Motion-compensated image reconstruction with alternating minimization. In *Proc. 2nd Intl. Mtg. on image formation in X-ray CT*, pages 330–333, 2012.

[26] S. Y. Chun. *Motion aspects in joint image reconstruction and nonrigid motion estimation*. PhD thesis, Univ. of Michigan, Ann Arbor, MI, 48109-2122, Ann Arbor, MI, 2008.

[27] S. Y. Chun and J. A. Fessler. Joint image reconstruction and nonrigid motion estimation with a simple penalty that encourages local invertibility. In *Proc. SPIE 7258 Medical Imaging 2009: Phys. Med. Im.*, page 72580U, 2009.

[28] S. Y. Chun and J. A. Fessler. A simple regularizer for B-spline nonrigid image registration that encourages local invertibility. *IEEE J. Sel. Top. Sig. Proc.*, 3(1):159–69, February 2009. Special Issue on Digital Image Processing Techniques for Oncology.

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

[30] A. M. Cormack. Representation of a function by its line integrals, with some radiological applications. *J. Appl. Phys.*, 34(9):2722–7, September 1963.

[31] M. Dawood, N. Lang, X. Jiang, and K. P. Schafers. Lung motion correction on respiratory gated 3-D PET/CT images. *IEEE Trans. Med. Imag.*, 25(4):476–85, April 2006.

[32] B. De Man and S. Basu. Distance-driven projection and backprojection in three dimensions. *Phys. Med. Biol.*, 49(11):2463–75, June 2004.

[33] B. De Man, S. Basu, N. Chandra, B. Dunham, P. Edic, M. Iatrou, S. McOlash, P. Sainath, C. Shaughnessy, B. Tower, and E. Williams. CatSim: a new computer assisted tomography simulation environment. In *Proc. SPIE 6510 Medical Imaging 2007: Phys. Med. Im.*, page 65102G, 2007.

[34] A. R. De Pierro and M. E. B. Yamagishi. Fast EM-like methods for maximum 'a posteriori' estimates in emission tomography. *IEEE Trans. Med. Imag.*, 20(4):280–8, April 2001.

[35] R. C. Fair. On the robust estimation of econometric models. *Ann. Econ. Social Measurement*, 2:667–77, October 1974.

[36] L. A. Feldkamp, L. C. Davis, and J. W. Kress. Practical cone beam algorithm. *J. Opt. Soc. Am. A*, 1(6):612–9, June 1984.

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

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

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

[40] J. A. Fessler. Statistical image reconstruction methods for transmission tomography. In M. Sonka and J. Michael Fitzpatrick, editors, *Handbook of Medical Imaging, Volume 2. Medical Image Processing and Analysis*, pages 1–70. SPIE, Bellingham, 2000.

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

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

[43] M. A. T. Figueiredo and José M Bioucas-Dias. Restoration of Poissonian images using alternating direction optimization. *IEEE Trans. Im. Proc.*, 19(12):3133–45, December 2010.

[44] D. Geman and G. Reynolds. Constrained restoration and the recovery of discontinuities. *IEEE Trans. Patt. Anal. Mach. Int.*, 14(3):367–83, March 1992.

[45] D. R. Gilland, B. A. Mair, J. E. Bowsher, and R. J. Jaszczak. Simultaneous reconstruction and motion estimation for gated cardiac ECT. *IEEE Trans. Nuc. Sci.*, 49(5):2344–9, October 2002.

[46] T. Goldstein and S. Osher. The split Bregman method for L1-regularized problems. *SIAM J. Imaging Sci.*, 2(2):323–43, 2009.

[47] M. Grass, R. Manzke, T. Nielsen, P. Koken, R. Proksa, M. Natanzon, and G. Shechter. Helical cardiac cone beam reconstruction using retrospective ECG gating. *Phys. Med. Biol.*, 48(18):3069–84, September 2003.

[48] E. J. Gravier and Y. Yang. Motion-compensated reconstruction of tomographic image sequences. *IEEE Trans. Nuc. Sci.*, 52(1):51–6, February 2005.

[49] C. Hamaker and D. C. Solmon. The angles between the null spaces of X rays. *J. Math. Anal. Applic.*, 62(1):1–22, January 1978.

[50] E. Hansis, D. Schäfer, O. Dössel, and M. Grass. Projection-based motion compensation for gated coronary artery reconstruction from rotational x-ray angiograms. *Phys. Med. Biol.*, 53(14):3807–20, July 2008.

[51] E. Hansis, H. Schomberg, K. Erhard, O. Dössel, and M. Grass. Four-dimensional cardiac reconstruction from rotational x-ray sequences: first results for 4D coronary angiography. In *Proc. SPIE 7258 Medical Imaging 2009: Phys. Med. Im.*, page 72580B, 2009.

[52] Eberhard Hansis, Dirk Schafer, Michael Grass, and Olaf Dossel. An iterative method for the reconstruction of the coronary arteries from rotational x-ray angiography. In *Proc. SPIE Medical Imaging 2007: Phys. Med. Im.*, 2007.

[53] W. R. Hendee and M. K. O'Connor. Radiation risks of medical imaging: Separating fact from fantasy. *Radiology*, 264(2):312–21, August 2012.

[54] G. T. Herman and L. B. Meyer. Algebraic reconstruction techniques can be made computationally efficient. *IEEE Trans. Med. Imag.*, 12(3):600–9, September 1993.

[55] J. Hinkle, M. Szegedi, B. Wang, B. Salter, and S. Joshi. 4D CT image reconstruction with diffeomorphic motion model. *Med. Im. Anal.*, 16(6):1307–16, August 2012.

[56] J. Hsieh. *Computed tomography: Principles, design, artifacts, and recent advances*. SPIE, Bellingham, 2003.

[57] J. Hsieh, O. E. Gurmen, and K. F. King. Investigation of a solid-state detector for advanced computed tomography. *IEEE Trans. Med. Imag.*, 19(9):930–40, September 2000.

[58] P. J. Huber. *Robust statistics*. Wiley, New York, 1981.

[59] H. M. Hudson and R. S. Larkin. Accelerated image reconstruction using ordered subsets of projection data. *IEEE Trans. Med. Imag.*, 13(4):601–9, December 1994.

[60] D. R. Hunter and K. Lange. A tutorial on MM algorithms. *American Statistician*, 58(1):30–7, February 2004.

[61] M. Iatrou, J. D. Pack, R. Bhagalia, D. Beque, and J. Seamans. Coronary artery motion estimation and compensation: a feasibility study. In *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, pages 2819–2821, 2010.

[62] K. Imai, M. Ikeda, Y. Enchi, and T. Niimi. Statistical characteristics of streak artifacts on CT images: Relationship between streak artifacts and mA s values. *Med. Phys.*, 36(2):492–9, February 2009.

[63] A. A. Isola, A. Ziegler, T. Koehler, W. J. Niessen, and M. Grass. Motion-compensated iterative cone-beam CT image reconstruction with adapted blobs as basis functions. *Phys. Med. Biol.*, 53(23):6777–98, December 2008.

[64] A. A. Isola, A. Ziegler, D. Schafer, T. Koehler, W. J. Niessen, and M. Grass. Motion-compensated iterative reconstruction of a region of interest in cardiac cone-beam CT. *Comput. Med. Imaging Graph.*, 34(2):149–159, March 2010.

[65] M. W. Jacobson and J. A. Fessler. An expanded theoretical treatment of iteration-dependent majorize-minimize algorithms. *IEEE Trans. Im. Proc.*, 16(10):2411–22, October 2007.

[66] Y. T. Jian and R. E. Carson. Feasible uniform-resolution penalized likelihood reconstruction for static- and multi-frame 3D PET. In *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, 2013.

[67] P. M. Joseph and R. D. Spital. A method for correcting bone induced artifacts in computed tomography scanners. *J. Comp. Assisted Tomo.*, 2(1):100–8, January 1978.

[68] M. Kachelriess and W. A. Kalender. Electrocardiogram-correlated image reconstruction from subsecond spiral computed tomography scans of the heart. *Med. Phys.*, 25(12):2417–31, December 1998.

[69] M. K. Kalra, M. M. Maher, M. A. Blake, B. C. Lucey, K. Karau, T. L. Toth, G. Avinash, E. F. Halpern, and S. Saini. Detection and characterization of lesions on low-radiation-dose abdominal CT images postprocessed with noise reduction filters. *Radiology*, 232:791–7, 2004.

[70] M. K. Kalra, M. M. Maher, S. Rizzo, and S. Saini. Radiation exposure and projected risks with multidetector-row computed tomography scanning: Clinical strategies and technologic developments for dose reduction. *J. Comp. Assisted Tomo.*, 28(S1):S46–9, July 2004.

[71] C. Kamphuis and F. J. Beekman. Accelerated iterative transmission CT reconstruction using an ordered subsets convex algorithm. *IEEE Trans. Med. Imag.*, 17(6):1001–5, December 1998.

[72] V. M. Kibardin. Decomposition into functions in the minimization problem. *Avtomatika i Telemekhanika*, 9:66–79, September 1979. Translation: p. 1311-23 in Plenum Publishing Co. "Adaptive Systems".

[73] D. Kim and J. A. Fessler. Ordered subsets acceleration using relaxed momentum for X-ray CT image reconstruction. In *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, pages 1–5, 2013.

[74] D. Kim, D. Pal, J-B. Thibault, and J. A. Fessler. Accelerating ordered subsets image reconstruction for X-ray CT using spatially non-uniform optimization transfer. *IEEE Trans. Med. Imag.*, 32(11):1965–78, November 2013.

[75] D. Kim, S. Ramani, and J. A. Fessler. Accelerating X-ray CT ordered subsets image reconstruction with Nesterov's first-order methods. In *Proc. Intl. Mtg. on Fully 3D Image Recon. in Rad. and Nuc. Med*, pages 22–5, 2013.

[76] H. Kudo, H. Nakazawa, and T. Saito. Convergent block-iterative method for general convex cost functions. In *Proc. Intl. Mtg. on Fully 3D Image Recon. in Rad. and Nuc. Med*, pages 247–250, 1999.

[77] P. J. La Riviere, J. Bian, and P. A. Vargas. Penalized-likelihood sinogram restoration for computed tomography. *IEEE Trans. Med. Imag.*, 25(8):1022–36, August 2006.

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

[79] Y. Long, J. A. Fessler, and J. M. Balter. 3D forward and back-projection for X-ray CT using separable footprints. *IEEE Trans. Med. Imag.*, 29(11):1839–50, November 2010.

[80] A. A. Mahabadi, S. Achenbach, C. Burgstahler, T. Dill, R. Fischbach, A. Knez, W. Moshage, B. M. Richartz, D. Ropers, S. Schröder, S. Silber, S. Möhlenkamp, et al. Safety, efficacy, and indications of β-adrenergic receptor blockade to reduce heart rate prior to coronary ct angiography. *Radiology*, 257(3):614–23, December 2010.

[81] F. A. Mettler, B. R. Thomadsen, M. Bhargavan, D. B. Gilley, J. E. Gray, J. A. Lipoti, J. McCrohan, T. T. Yoshizumi, and M. Mahesh. Medical radiation exposure in the US in 2006: Preliminary results. *Health Physics*, 95(5):502–7, November 2008.

[82] K. Muller, Y. Zhang, G. Lauritsch, C. Rohkohl, C. Schwemmer, A. Maier, R. Fahrig, and J. Hornegger. Evaluation of interpolation methods for motion compensated tomographic reconstruction for cardiac angiographic c-arm data. In *Proc. 2nd Intl. Mtg. on image formation in X-ray CT*, pages 5–8, 2012.

[83] R. J. Murphy, S. Yan, J. A. O'Sullivan, D. L. Snyder, B. R. Whitling, D. G. Politte, G. Lasio, and J. F. Williamson. Pose estimation of known objects during transmission tomographic image reconstruction. *IEEE Trans. Med. Imag.*, 25(10):1392–404, October 2006.

[84] S. Mustafovic and K. Thielemans. Object dependency of resolution in reconstruction algorithms with inter-iteration filtering applied to PET data. *IEEE Trans. Med. Imag.*, 23(4):433–46, April 2004.

[85] O. Nalcioglu and R. Y. Lou. Post-reconstruction method for beam hardening in computerised tomography. *Phys. Med. Biol.*, 24(2):330–40, March 1979.

[86] R. Neal and G. E. Hinton. A view of the EM algorithm that justifies incremental, sparse and other variants. In M. I. Jordan, editor, *Learning in Graphical Models*, pages 255–68. Kluwer, Dordrencht, 1998.

[87] H. Nien and J. A. Fessler. Combining augmented Lagrangian method with ordered subsets for X-ray CT image reconstruction. In *Proc. Intl. Mtg. on Fully 3D Image Recon. in Rad. and Nuc. Med*, pages 280–3, 2013.

[88] J. Nocedal. Large scale unconstrained optimization. In A. Watson and I. Duff, editors, *The State of the Art in Numerical Analysis*. Oxford University Press, Oxford, 1996.

[89] J. Nocedal and S. J. Wright. *Numerical optimization*. Springer, New York, 1999.

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

[91] J. M. Ollinger and J. A. Fessler. Positron emission tomography. *IEEE Sig. Proc. Mag.*, 14(1):43–55, January 1997.

[92] S. S. Orlov. Theory of three dimensional reconstruction. I. Conditions for a complete set of projections. *Sov. Phys. Crystallography*, 20:312–14, 1976.

[93] X. Pan, L. Yu, and C-M. Kao. Spatial-resolution enhancement in computed tomography. *IEEE Trans. Med. Imag.*, 24(2):246–53, February 2005.

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

[95] D. L. Parker. Optimal short scan convolution reconstruction for fan beam CT. *Med. Phys.*, 9(2):254–7, March 1982.

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

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

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

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

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

[101] S. Ramani and J. A. Fessler. An augmented Lagrangian method for regularized MRI reconstruction using SENSE. In *Proc. Intl. Soc. Mag. Res. Med.*, page 2873, 2011.

[102] S. Ramani and J. A. Fessler. Convergent iterative CT reconstruction with sparsity-based regularization. In *Proc. Intl. Mtg. on Fully 3D Image Recon. in Rad. and Nuc. Med*, pages 302–5, 2011.

[103] S. Ramani and J. A. Fessler. Parallel MR image reconstruction using augmented Lagrangian methods. *IEEE Trans. Med. Imag.*, 30(3):694–706, March 2011.

[104] S. Ramani and J. A. Fessler. Regularized parallel MRI reconstruction using an alternating direction method of multipliers. In *Proc. IEEE Intl. Symp. Biomed. Imag.*, pages 385–8, 2011.

[105] S. Ramani and J. A. Fessler. A splitting-based iterative algorithm for accelerated statistical X-ray CT reconstruction. *IEEE Trans. Med. Imag.*, 31(3):677–88, March 2012.

[106] C. Rohkohl, H. Bruder, K. Stierstorfer, and T. Flohr. Improving best-phase image quality in cardiac CT by motion correction with MA optimization. In *Proc. 2nd Intl. Mtg. on image formation in X-ray CT*, pages 1–4, 2012.

[107] K. Sauer and C. Bouman. A local update strategy for iterative reconstruction from projections. *IEEE Trans. Sig. Proc.*, 41(2):534–48, February 1993.

[108] K. Sauer, C. Bouman, and J. B. Thibault. Noise modeling for model-based image reconstruction in X-ray CT, 2014. Preprint.

[109] D. Schafer, J. Borgert, V. Rasche, and M. Grass. Motion-compensated and gated cone beam filtered back-projection for 3-D rotational X-ray angiography. *IEEE Trans. Med. Imag.*, 25(7):898–906, July 2006.

[110] S. Schmitt and J. A. Fessler. Fast variance computation for quadratically penalized iterative reconstruction of 3D axial CT images. In *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, pages 3287–92, 2012.

[111] U. J. Schoepf, C. R. Becker, B. M. Ohnesorge, and E. K. Yucel. CT of coronary artery disease. *Radiology*, 232(1):18–37, July 2004.

[112] W. P. Segars, M. Mahesh, T. J. Beck, E. C. Frey, and B. M. W. Tsui. Realistic CT simulation using the 4D XCAT phantom. *Med. Phys.*, 35(8):3800–8, August 2008.

[113] H. Shi. *Fast regularization design for tomographic image reconstruction for uniform and isotropic spatial resolution*. PhD thesis, Univ. of Michigan, Ann Arbor, MI, 48109-2122, Ann Arbor, MI, 2008.

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

[115] H. Shi and J. A. Fessler. Quadratic regularization design for 3D axial CT. In *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, pages 2834–6, 2006.

[116] H. R. Shi and J. A. Fessler. Quadratic regularization design for 2D CT. *IEEE Trans. Med. Imag.*, 28(5):645–56, May 2009.

[117] S. Sotthivirat and J. A. Fessler. Relaxed ordered-subsets algorithm for penalized-likelihood image restoration. *J. Opt. Soc. Am. A*, 20(3):439–49, March 2003.

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

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

[120] K. Taguchi, B. S. Chiang, and I. A. Hein. Direct cone-beam cardiac reconstruction algorithm with cardiac banding artifact correction. *Med. Phys.*, 33(2):521–39, February 2006.

[121] K. Taguchi and H. Kudo. Motion compensated fan-beam reconstruction for nonrigid transformation. *IEEE Trans. Med. Imag.*, 27(7):907–17, July 2008.

[122] Q. Tang, J. Cammin, S. Srivastava, and K. Taguchi. A fully four-dimensional, iterative motion estimation and compensation method for cardiac CT. *Med. Phys.*, 39(7):4291–305, July 2012.

[123] Q. Tang, J. Cammin, and K. Taguchi. Four-dimensional projection-based motion estimation and compensation for cardiac x-ray computed tomography. In *Proc. Intl. Mtg. on Fully 3D Image Recon. in Rad. and Nuc. Med*, pages 46–9, 2013.

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

[125] J. H. Thrall. Radiation exposure in CT scanning and risk: where are we? *Radiology*, 264(2):325–8, August 2012.

[126] H. Tong, M. Li, H.-J. Zhang, and C. Zhang. Blur detection for digital images using wavelet transform. In *Multimedia and Expo, 2004. ICME '04. 2004 IEEE International Conference on*, volume 1, pages 17–21, June 2004.

[127] H. K. Tuy. An inversion formula for cone-beam reconstruction. *SIAM J. Appl. Math.*, 43(3):546–52, June 1983.

[128] M. Unser, A. Aldroubi, and M. Eden. B-spline signal processing: Part I—theory. *IEEE Trans. Sig. Proc.*, 41(2):821–33, February 1993.

[129] U. van Stevendaal, J. von Berg, C. Lorenz, and M. Grass. A motion-compensated scheme for helical cone-beam reconstruction in cardiac CT angiography. *Med. Phys.*, 35(7):3239–51, July 2008.

[130] N. Villain, Y. Goussard, J. Idier, and M. Allain. Three-dimensional edge-preserving image enhancement for computed tomography. *IEEE Trans. Med. Imag.*, 22(10):1275–87, October 2003.

[131] J. Wang, T. Li, H. Lu, and Z. Liang. Penalized weighted least-squares approach to sinogram noise reduction and image reconstruction for low-dose X-ray computed tomography. *IEEE Trans. Med. Imag.*, 25(10):1272–83, October 2006.

[132] B. R. Whiting, P. Massoumzadeh, O. A. Earl, J. A. O'Sullivan, D. L. Snyder, and J. F. Williamson. Properties of preprocessed sinogram data in x-ray computed tomography. *Med. Phys.*, 33(9):3290–303, September 2006.

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

[134] Q. Xu, H. Yu, X. Mou, L. Zhang, J. Hsieh, and G. Wang. Low-dose X-ray CT reconstruction via dictionary learning. *IEEE Trans. Med. Imag.*, 31(9):1682–97, September 2012.

[135] W. Xu and K. Mueller. Evaluating popular non-linear image processing filters for their use in regularized iterative CT. In *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, pages 2864–5, 2010.

[136] Y. Xu, W. He, H. Chen, Z. Hu, J. Li, and T. Zhang. Impact of the adaptive statistical iterative reconstruction technique on image quality in ultra-low-dose CT. *Clinical Radiology*, 68(9):902–8, September 2013.

[137] L. Yu, X. Liu, S. Leng, J. M. Kofler, J. C. Ramirez-Giraldo, M. Qu, J. Christner, J. G. Fletcher, and C. H. McCollough. Radiation dose reduction in computed tomography: techniques and future perspective. *Imaging in Medicine*, 1(1):65–84, October 2009.

[138] Z. Yu, C. A. Bouman, J-B. Thibault, and K. D. Sauer. Image grid invariant regularization for iterative reconstruction. In *Proc. Intl. Mtg. on Fully 3D Image Recon. in Rad. and Nuc. Med*, pages 517–20, 2013.

[139] Z. Yu, L. Fu, D. Pal, J-B. Thibault, C. A. Bouman, and K. D. Sauer. Nested loop algorithm for parallel model based iterative reconstruction. In *Proc. Intl. Mtg. on Fully 3D Image Recon. in Rad. and Nuc. Med*, pages 197–200, 2013.

[140] Z. Yu, J-B. Thibault, C. A. Bouman, K. D. Sauer, and J. Hsieh. Fast model-based X-ray CT reconstruction using spatially non-homogeneous ICD optimization. *IEEE Trans. Im. Proc.*, 20(1):161–75, January 2011.

[141] G. Zeng. A filtered backprojection MAP algorithm with nonuniform sampling and noise modeling. *Med. Phys.*, 39(4):2170–8, April 2012.

[142] G. L. Zeng and A. Zamyatin. A filtered backprojection algorithm with ray-by-ray noise weighting. *Med. Phys.*, 40(3):031113, March 2013.

[143] K. Zeng, B. De Man, J-B. Thibault, Z. Yu, C. Bouman, and K. Sauer. Spatial resolution enhancement in CT iterative reconstruction. In *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, pages 3748–51, 2009.

[144] H. Zhang, J. Ma, J. Wang, Y. Liu, H. Lu, and Z. Liang. Statistical image reconstruction for low-dose CT using nonlocal means-based regularization. *Computerized Medical Imaging and Graphics*, 2014.

[145] A. Ziegler, T. Nielsen, and M. Grass. Iterative reconstruction of a region of interest for transmission tomography. *Med. Phys.*, 35(4):1317–27, April 2008.