Analysis of Compressed Sensing Based Computed Tomography Reconstruction

Faculty of Science, Engineering and Technology
Swinburne University of Technology
Melbourne, Australia

Submitted for the degree of Doctor of Philosophy

Wen Hou

2015

Abstract

This thesis studies the application of compressed sensing (CS) to computed tomography (CT) for high quality CT image reconstruction with low dose X-ray radiation. X-ray CT has been a popular medical imaging modality because of its ability to provide a wide coverage of the examination area, fast scanning speed, high spatial resolution, ease of operation and relatively low operation cost. It plays an important role in diagnosis and thus benefits patients greatly. In general, a considerable amount of X-ray radiation is needed for the production of CT images of high quality, but increased X-ray radiation dose can damage healthy cells and raise the risk of cancer. Therefore, to reduce the X-ray radiation dose to patients and preserve imaging quality has been a significant and challenging problem in CT development. CS technology is an emerging research area which recovers signals from far fewer samples than required by the Shannon-Nyquist sampling theorem. There have been a number of CS-based techniques applied to CT reconstruction problems subjected to under-sampled or noisy data.

The theoretical results from CS do not extend to the CT setting. There is a fundamental lack of understanding about which type of under-sampling is favourable to CS reconstruction of CT images. One reason is that the CS theory requires the sampling matrix in the sparsifying domain to meet an incoherence condition in terms of the restricted isometry property (RIP), which cannot be tested by a known polynomial-time algorithm. It is a key challenge as well as the first task of this thesis to carry out analysis of the incoherence property of the CT data scanning schemes for successful CS reconstruction. The property of the system matrix of fan-beam CT is investigated to achieve low radiation dose while maintaining good reconstruction quality through CS. To reduce the radiation dose, scanning data is under-sampled in both the view and bin direction. For the limited-angle scanning, two sampling patterns are adopted: golden-angle and random-angle. For the sparse-bin setting, the reduced detectors are either evenly or randomly distributed. The analysis is conducted based on the point spread function (PSF) and the Fourier slice theorem (FST), respectively. The results of the analysis are verified by simulations and it is shown that high-quality CT reconstruction can

be achieved with random-detector and golden-angle scanning schemes.

The thesis further proposes two novel reconstruction models for the reduction of X-ray radiation dose in CT examinations without compromising image quality by CS. Theoretical analysis shows that both models conform to the requirement of CS, and successful reconstruction is validated by simulations. In the first model, the CT projections are first converted to Mojette projections. Then 1-dimensional Fourier transform is applied to the modified Mojette projections, followed by an exact mapping of the gained Fourier coefficients to the 2-dimensional Fourier domain of the scanning object. Finally the scanned object can be reconstructed from the partial Fourier coefficients by CS. In the second model, the smoothness of the sinogram data set and the sparsity of its frequency transformation are exploited, which in conjunction with a randomly projected X-ray radiation scheme, can result in a randomly under-sampled partial Fourier matrix as the sensing matrix for CS reconstruction. The formulation of this CT data acquisition and reconstruction scheme satisfies the incoherence and sparsity properties required by CS theory. Based on this scheme, a weighted ℓ_1 regularized optimization algorithm is proposed for computing the CS image reconstruction. The reconstruction performance and advantages over other known CT reconstruction methods are demonstrated using simulated phantom and CT images.

To summarize, the contributions of the thesis are two-fold. First, we bridge the gap between the theoretical CS requirement and the practical CT setting. Secondly, two novel models for CT reconstruction are proposed, which apply the CS theory to produce satisfactory CT images.

Dedicated to my mother, Donglan Zhang

Acknowledgements

I would like to express my deep gratitude to my research supervisor Professor Cishen Zhang, for his patient guidance, enthusiastic encouragement and useful critiques of this research work. His constant kindness is especially appreciated. In addition, I would like to thank Professor Zhihong Man for his co-supervison. Thanks are also due to the various developers of open source software systems for releasing their software with non-restrictive licensing. I am grateful to my current employer, Swinburne University of Technology, for providing the resources and support to pursue the doctoral degree. Dr. Alex McKnight assisted by proofreading the final draft for grammar and style.

Finally, I would like to thank my family and friends for their loving forbearance during the long period it has taken me to conduct the research and write up this thesis.

Wen Hou, 2015

Declaration

I declare that this thesis contains no material that has been accepted for the award of any other degree or diploma and to the best of my knowledge contains no material previously published or written by another person except where due reference is made in the text of this thesis.

Wen Hou, 2015

Contents

List of Figures

List of Tables

Symbols

N_{in}	total number of photons entering the object
N_{out}	total number of photons exiting the object
$\mu(x, y)$	attenuation function of the object with two space coordinates x and y
$S_{in}(E)$	incident photon number density
$S_{out}(E)$	exiting photon number density
$f(x, y)$	object image function at the 2-D rectangular coordinate system (x, y)
$s(\theta, r)$	parallel-beam CT projections with the angular and radial coordinates (θ, r)
$s_f(\beta, \gamma)$	fan-beam CT projections with the rotation angle β and γ specifies the ray location within a fan
$\delta(\cdot)$	Dirac delta function
$\hat{S}(\theta, \rho)$, $\hat{S}_c(\theta, \rho)$	1-D CTFT of $s(\theta, r)$ with regard to r
$\hat{S}_d(\theta, \rho)$	1-D DTFT of $s(\theta, r)$ with regard to r
R	radius of the source trajectory in fan-beam CT
D	radius of the curved detector array centered at the source in fan-beam CT
$n \times n$	image size
A	measurement/sensing matrix of size $m \times N$
$\mathbf{x}$	vectorized object signal of size $N \times 1$
$\mathbf{b}$	measurement vector of size $m \times 1$
$\mathbb{C}$	set of complex numbers
$\mathbb{R}$	set of real numbers
$\|\|\mathbf{x}\|\|_p$	ℓ_p norm, defined as $\|\|\mathbf{x}\|\|_p = (\sum_{\ell=1}^{N} \|\mathbf{x}_\ell\|^p)^{1/p}, 1 \leq p < \infty$
$\|\|\mathbf{x}\|\|_0$	ℓ_0 norm, defined as the number of nonzero entries of $\mathbf{x}$
$\|\|\mathbf{x}\|\|_1$	ℓ_1 norm, defined as the sum of the absolute entries of $\mathbf{x}$
$\|\|\mathbf{x}\|\|_\infty$	maximum norm, defined as the maximum element of $\mathbf{x}$
σ, ϵ, α, η	parameters
$\|\cdot\|$	cardinality
$\text{supp}(\mathbf{x})$	support of $\mathbf{x}$
k	sparsity, i.e. there are at most k non-zero entries
φ	orthonormal basis

$\mathcal{N}(A)$	null space of the measurement matrix A
$\sum_k$	set of k-sparse vectors
δ_k	restricted isometry constant
F	discrete Fourier matrix
F^*, F_1^{-1}	inverse discrete Fourier matrix
$\hat{F}(u,v)$, $\hat{F}_c(u,v)$	2-D CTFT of $f(x,y)$ with u,v as the index in frequency domain
$\hat{F}_d(u,v)$	2-D DTFT of $f(x,y)$
I	identity matrix
$g_c(t)$	continuous-time signal
$G_c(\Omega)$	CTFT of $g_c(t)$
$g_d[l]$	discrete-time version of $g_c(t)$, i.e. $g_d[l] = g_c(lT)$ with T as the sampling period
$G_d(w)$	DTFT of $g_d[l]$
F_2	2-D Fourier transform matrix
Δ_β and Δ_γ	step of β and γ
$\|\|A\|\|_2$	ℓ_2 norm of A, defined as $\|\|A\|\|_2 = \sigma_{\max}(A)$
$\|\|A\|\|_F$	Frobenius norm of $A = (a_{jk}) \in \mathbb{C}^{m\times N}$, defined as $\|\|A\|\|_F = \sqrt{\sum_j \sum_k \|a_{jk}\|^2}$
$[N]$	$[N] = 1, 2, \cdots, N$
$\mathbf{K}$	$\mathbf{K} \subset [N], \|\mathbf{K}\| \leq k$
$A_{\mathbf{K}}$	column sub-matrix of A consisting of the columns indexed by $\mathbf{K}$
Υ	quantified index of RIP, defined as $\|\|A^*A - I\|\|_F$
μ	current quantified index of RIP, defined as $\max(A^*A - I)$
$\mathcal{F}(u,v)$	2-D DFT of $f(x,y)$
$M_{p,q}(r)$	Dirac-Mojette projections over angles $\theta = \tan^{-1}(\frac{q}{p})$ where p and q are relatively prime integers
$\mathcal{FM}_{p,q}(\rho)$	1-D DFT of $M_{p,q}(r)$
$M'_{p,q}(r)$	modified Mojette projections
$\mathcal{FM}'_{p,q}(\rho)$	1-D DFT of $M'_{p,q}(r)$
$\mathbf{s}_r$	vectorized representation of the sinogram $s(\theta, r)$
$\hat{\mathbf{s}}_r$	DFT of $\mathbf{s}_r$
$\tilde{\mathbf{s}}_r$	under-sampled sinogram data vector
$\mathbf{\Phi}_r$	partial inverse Fourier matrix
$\circ$	component-wise multiplication
$\text{sgn}(z)$	sign function, defined as $\begin{cases} -1 & \text{if } z < 0 \\ 0 & \text{if } z = 0 \\ 1 & \text{if } z > 0 \end{cases}$

Abbreviations

CS	compressed/compressive sensing
CT	computed tomography
AEC	automatic exposure control
ROI	region of interest
FFT	fast Fourier transform
FST	Fourier slice theorem
OMP	orthogonal matching pursuit
RIP	restricted isometry property
ART	algebraic reconstruction technique
SART	simultaneous algebraic reconstruction technique
OS	ordered subsets
MART	multiplicative algebraic reconstruction technique
ML-EM	maximum likelihood expectation-maximization
LS	least squares
TV	total variation
CTFT	continuous-time Fourier transform
DTFT	discrete-time Fourier transform
PICCS	prior image constrained compressed sensing
GPU	graphics processing units
HU	Hounsfield units
SPR	sidelobe-to-peak ratio

Chapter 1

Introduction

1.1 Motivation and Objective

X-ray computed tomography (CT) is a technology that provides the user with virtual slices inside the object by using computer-processed X-rays to produce tomographic images of specific areas of a scanned object instead of cutting the object. The inner structures are revealed based on their ability to attenuate the X-ray beam. The significant clinical benefits of CT in medical imaging and health care were immediately recognized following its introduction into clinical practice in 1972, when Sir Godfrey Hounseld at EMI patented the first CT scanner. Since that time, there has been extensive growth in the number of CT scanners and the frequency of CT examinations. It was cited as one of the most significant medical innovations in the previous decades in a US survey [121]. The use of CT scans in US has increased more than 3-fold since 1993 to approximately 70 million scans annually [11]. Its popularity is based on its capability of providing a wide coverage of the area of interest in a short time, high spatial resolution, ease of operation and relatively low acquisition cost. Hence, as a popular and useful medical tool, CT plays an important role in diagnosis and thus benefits patients greatly.

Despite the fact that CT provides invaluable information for diagnosis and patient management, there is concern about potential future cancer risks because CT scans involve much higher radiation doses than conventional diagnostic X-rays. It is known that the ionizing radiation can break important biomolecules

(e.g. DNA), damaging or killing the affected cell, or in the worst case causing cancer. CT alone contributes almost one half of the total radiation exposure from medical use and one quarter of the average radiation exposure per capita in US [1], but this high figure is mostly due to the large number of persons exposed annually. An Australian study of 10.9 million people [102] reported that the increased incidence of cancer after CT scan exposure in this cohort was mostly due to irradiation, where one in every 1800 CT scans was followed by an excess cancer. Hence, the risks to individuals are likely to be almost invisible compared to the health benefits of a timely, accurate and non-invasive diagnosis that CT facilitates. Nevertheless, since the cancer risk associated with the radiation dose in CT is not zero, the dose-conscious CT community will continue to list radiation reduction as one of the top priorities, particularly in light of the continued increase in the number of CT examinations performed annually worldwide [164].

The most intuitive and straightforward method of reducing radiation dose is to take fewer projections, which is the focus of this thesis. The reduction of projection data can also lead to other benefits, such as reducing scan time and improving time-resolution in cardiac CT. However, it creates challenges for maintaining a high image quality, as the projection data is not sufficient for exact reconstruction of tomographic images according to the Nyquist-Shannon sampling theorem. This fundamental theorem in signal processing expresses the sample rate in terms of the signal's bandwidth and states that a band-limited analogue signal that has been sampled can be perfectly reconstructed from an infinite sequence of samples if the sampling rate exceeds twice the signal's bandwidth. The rate is known as the Nyquist rate. When the projections are fewer for the sake of low radiation, the sample rate falls below the Nyquist rate and becomes so insufficient that the application of standard analytic algorithms such as filtered back-projection (FBP) can cause conspicuous artifacts in reconstructed images. Hence, the image quality of CT is generally proportional to the radiation dose, and to reduce the X-ray radiation dose to patients without compromising the imaging quality is an important and challenging problem in CT development.

Fortunately, the Nyquist-Shannon sampling theorem does not preclude the possibility of perfect reconstruction under special circumstances that do not satisfy the sample-rate criterion. The emerging technology, compressed sensing (CS), also known as compressive sampling or compressive sensing, can achieve perfect signal recovery from much fewer measurements than usually required by the clas-

sic Shannon-Nyquist criterion. There are however, two conditions to be satisfied for perfect recovery: the signal sparsity, and the restricted isometry property (RIP) of the sensing matrix. Basically, CS exploits the sparseness or compressibility of signals in a predefined basis/frame, and uses it as the constraint to solve an under-determined linear system. The required number of measurements is proportional to the number of non-zero elements in the sparse representation of signals. CS has gained wide attention since it was proposed in 2004 by Emmanuel Candès, Terence Tao, and David Donoho [21–23, 40, 149], and medical imaging is a prominent application for two obvious reasons. One is that most medical images can be sparsely represented by some linear transformation, and the other is that medical scanners obtain encoded samples. The most active area of CS application, however, is not in CT, but in magnetic resonance imaging (MRI) [26,94,95]. CS addresses the issue of high scan time in MRI by measuring fewer Fourier coefficients, and since the partial Fourier matrix has been proved to satisfy RIP, MRI reconstruction satisfies the requirement of CS. Hence, all the CS-based research in MRI has a solid theoretical foundation. In the case of CT, a number of CS-based techniques have also been introduced to CT reconstruction problems subject to under-sampled or noisy data [29, 70, 139, 140]. In spite of all these efforts, many problems are still open and call for further studies. One of the main challenges lies in the fact that the theoretical results from CS have not to date been extended to the CT setting. In other words, the CT system matrix has not been shown to satisfy RIP. There is a fundamental lack of understanding about which type of under-sampling is favourable to post CS reconstruction. In addition, reconstruction accuracy and speed can both be improved.

In summary, the application of CS has the advantage of radiation reduction while maintaining high recovery quality. It has therefore become one of the central topics in CT imaging. Our main object is to address the current disadvantages of CS in CT mentioned above and provide some more advanced understanding about the RIP of the CT system matrix, i.e. to achieve incoherence in designing CT data acquisition schemes for successful CS reconstruction. It may well be seen as a step forward in attempting to bridge the theoretical gap. We also expect to deepen our understanding by developing novel reconstruction models that not only conform to the requirement of CS, but also outperform existing methods with respect to accuracy and efficiency. There are three types of CT in terms of the beam configuration: parallel-beam, fan-beam, and cone-beam, with the former two types in 2-D domain and the third type in 3-D domain.

Despite the popularity of cone-beam CT in current diagnosis system, it involves complicated geometry settings as well as large-volume data, which makes the reconstruction quite computational demanding. To start with a relatively simple configuration, we make the analysis mostly based on fan-beam CT. As a matter of fact, cone-beam CT can be seen as a generation of fan-beam CT and these two configurations have many similarities. Since the fundamental CS-based principles remain the same, the fan-beam CT based-analysis in this thesis can provide some preliminary knowledge for the future analysis of cone-beam CT and the extension is quite straightforward.

1.2 Main Contributions

The main contributions of the thesis are as follows:

- In order to extend the theoretical results from CS to the CT setting, we decompose the system matrix based on Fourier slice theorem (FST), which is inspired by the fact that the RIP of the Fourier matrix is well defined. To the best of our knowledge, we are the first to tackle the RIP of the CT system matrix from this aspect. Four specific sampling schemes are studied: golden-angle and random-angle, and regular-detector and random-detector. After analyzing the effect of different schemes on the FST-based model, we reveal the favourable sampling schemes for CS-based reconstruction. This part of content can be mainly found in Chapters 4 and 5.
- With the same purpose as above, we take the analysis from another view based on point spread function (PSF). The Frobenius norm of the difference matrix between PSF and the identity matrix is proposed as an evaluation index. The lower the Frobenius norm, the better the reconstruction. In contrast, high off-diagonal interference in PSF implies low-quality recovery. Simulation results show that the index developed in this thesis is more sensitive than the maximum of the sidelobe-to-peak ratio (SPR), which is currently used to evaluate the incoherence of the system matrix. This part of content can be mainly found in Chapter 5.
- We gain deeper understanding of the direct application of CS in CT reconstruction by presenting a novel frame that satisfies the requirement of

CS for parallel-beam CT. The sinogram is first converted to the projections gained through Mojette transform, which is an exact discretization of the Radon transform. On each view angle, the projections are summed up with the application of certain principles. Then the 1-D Fourier coefficients of the merged projections are mapped to the 2-D Fourier domain of the object. Finally, compressed sensing is employed to deal with the partial Fourier coefficients and is shown to recover the object very well and suppress the small noise effectively. Experimental results have demonstrated the advantages of the proposed method. With the use of Mojette transform and compressed sensing, the purpose of reducing the radiation dosage during CT examinations without compromising image quality is achieved. This part of content can be mainly found in Chapter 6.

- CS is applied to indirectly reconstruct the sparse or compressible Fourier coefficients of the sinogram rather than to recover the object directly from the sinogram data set. The smoothness of the sinogram data set and the sparsity of its frequency transformation are exploited, which in conjunction with a randomly projected X-ray radiation scheme, can result in a randomly under-sampled partial Fourier matrix as the sensing matrix for CS reconstruction. The formulation of this CT data acquisition and reconstruction scheme satisfies the incoherence and sparsity properties required by CS theory. Based on this scheme, a weighted ℓ_1 regularized optimization algorithm is proposed for computing the CS image reconstruction. The reconstruction performance and advantages over other known CT reconstruction methods are demonstrated by simulated phantom and CT images. This part of content can be mainly found in Chapter 7.

1.3 Outline of Thesis

This thesis explores the application of CS in CT for low-dose radiation yet high-quality reconstruction, in an attempt to bridge the gap between the CS theory and the practical CT setting. The rest of the thesis is organized as follows:

In Chapter 2, the fundamentals of computed tomography are introduced. We start with the development of CT, from the very first scanner to different generations: parallel-beam, fan-beam and cone-beam. The physical theory of CT

is then described, as well as the cause and appearance of various artifacts. In the following section, we focus on low-radiation CT, revealing its benefits and guiding principles, the quantification of radiation and effective dose-management strategies. Finally, both the analytical and iterative reconstruction algorithms are depicted in detail, together with their advantages and disadvantages.

In Chapter 3, the foundation of compressed sensing is laid. To begin with, we provide the problem statement and give the associated two main solvers: basis pursuit and greedy algorithms. Then the two key elements of CS theorem are described in detail: sparse/compressible signals and the RIP of the sensing matrix. These two work together to guarantee the uniqueness of the solution. RIP for a random Fourier matrix is specifically mentioned as it is related closely to the CT system as will be seen in Chapter 4. Finally, we present the application of CS in CT, which highlights the TV regularization and the prior image constrained compressed sensing (PICSS).

In Chapter 4, we analyse sampling schemes in CT. It is a transitional chapter, which prepares for the FST-based analysis of the system matrix in Chapter 5. First, the full scanning fan-beam CT is illustrated, followed by the introduction of the sparse-view and sparse-detector CT, which leads to four specific sampling schemes: golden-angle and random-angle, and regular-detector and random-detector. Then we build the projection model of parallel-beam CT based on the preliminary knowledge of FST and the relation between different Fourier transforms. The analysis of the model states that the error becomes negligible when the padding size is sufficiently large. Next, we bridge the gap between the built model and fan-beam CT by revealing the rebinning relationship between fan-beam CT and parallel-beam CT, and depict in detail what the model is like for full-scanning, sparse-view and sparse-detector fan-beam CT, respectively. Finally, a simple simulation is run to validate the accuracy of the model compared with the Matlab built-in model and the linear intersection model.

In Chapter 5, the property of the system matrix for fan-beam CT is investigated with the purpose of both reducing the radiation dose and preserving high-quality recovery. With under-sampled projections, CS is employed for recovery using the optimization algorithm of TVAL3, and has shown advantages over FBP in the simulations. Four under-sampling patterns are considered: golden-angle and random-angle, and regular-detector and random-detector. Two tools are used for

the analysis of the impacts of different sampling schemes on reconstructions: PSF and FST. Based on PSF, an evaluation index is proposed: the Frobenius norm Υ of the difference matrix between PSF and the identity matrix. The lower the Υ, the better the reconstruction. In contrast, high off-diagonal interference in PSF implies low-quality recovery. Based on FST, the system matrix is decomposed to Fourier matrix and our aim is converted to the study of sampling on the Fourier matrix. It is shown that the golden-angle system obtains lower Υ than the random-angle and the random-detector setting obtains lower Υ than the regular-detector. Experiments are conducted on phantom and real CT images, for which the results are measured by normalized mean square error (NMSE). The results are consistent with the indicator results, with golden-angle scanning and random-detector setting obtaining lower NMSE than the other two, respectively. With both the theoretical and experimental analysis, the purpose of reducing radiation dose without compromising the reconstruction quality is achieved. The conclusion is that golden-angle and random-detector scanning is the most favourable for the post CS-based reconstruction.

In Chapter 6, we analyse the direct application of CS in CT by presenting a novel frame for parallel-beam CT reconstruction. First, the sinogram is converted to the projections gained through Mojette transform, which is an exact discretization of the Radon transform. On each view angle, the projections are summed up with the application of certain principles. Then the 1-D Fourier coefficients of the merged projections are mapped to the 2-D Fourier domain of the object. Finally, compressed sensing is employed to deal with the partial Fourier coefficients via the specific optimization algorithm of RecPF, which can recover the object very well and suppress the small noise effectively. Experimental results have demonstrated the advantages of the proposed method. With the use of Mojette transform and compressed sensing, the purpose of reducing the radiation dosage during CT examinations without compromising image quality is achieved.

In Chapter 7, we analyse the indirect application of CS in CT reconstruction. A CS approach to fan-beam CT image reconstruction is proposed for the purpose of reducing the X-ray radiation dose in CT examinations without compromising image quality. The proposed approach is novel in the sense that it overcomes the difficulties of the existing CS approaches to CT image reconstruction which have not been able to establish the incoherence and sparsity conditions required by the CS theory for signal reconstruction. Instead of applying CS to the reconstruc-

tion of the object directly from the sinogram data set, the proposed CS approach indirectly reconstructs the sparse or compressible DFT of the sinogram via the optimization algorithm of yall1. The correspondingly formulated sensing scheme results in a randomly sampled partial DFT matrix which meets the theoretical RIP condition for CS reconstruction. The proposed weighted ℓ_1 regularized optimization algorithm takes into account the low pass property of the reconstruction signal and its reconstruction results outperform the existing CSTV reconstruction method.

In Chapter 8, we conclude the thesis and highlight some potential future research directions.

Chapter 2

Fundamentals of Computed Tomography

2.1 History

X-ray computed tomography (CT) is a non-invasive technology that combines a series of X-ray projections and computer processing methods to produce tomographic images (virtual 'slices') of specific areas of the scanned object. The most common application of X-ray CT is in medical imaging, as it provides a unique way to look at the interior of a patient for diagnostic and therapeutic purposes. Compared to traditional 2-D medical radiography, CT completely eliminates the superimposition of images of structures outside the region of interest (ROI). It possesses inherent high-contrast resolution and thus can distinguish differences between tissues that differ in physical density by less than 1%. In addition, CT is a multi-planar reformatted imaging technique and the scanned data can be viewed as images in the transverse, coronal, or sagittal planes, as shown in Figure 2.1.

The history of the medical CT scanner is briefly recapitulated here [78] and illustrated in Figure 2.2. The first commercially viable CT scanner was invented by Sir Godfrey Hounsfield in the early 1970s, and is known as the first-generation system. It used a parallel-beam/pencil-beam geometry and the projections were measured by moving an X-ray source and detector along parallel lines on opposite sides of an object. As it could record only a single data point at a time, this system was very time-consuming and soon replaced by the second generation scanners

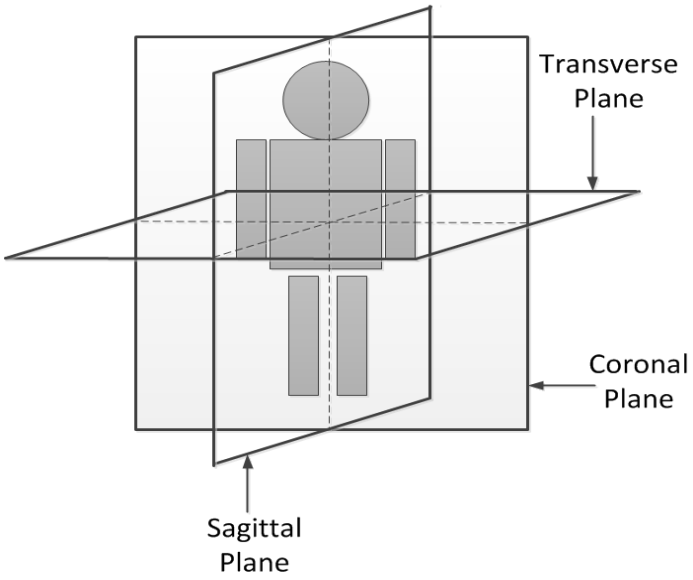
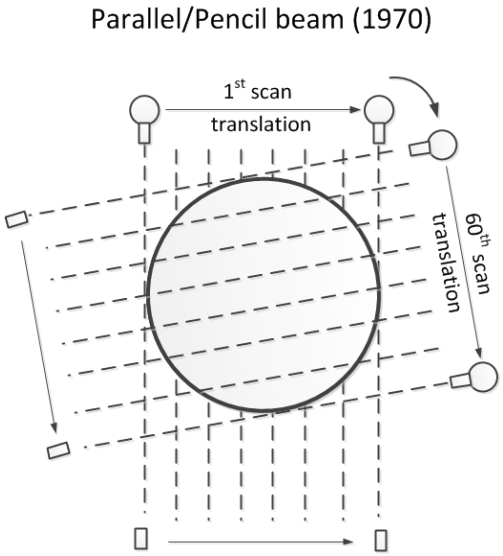


Figure 2.1: Body planes

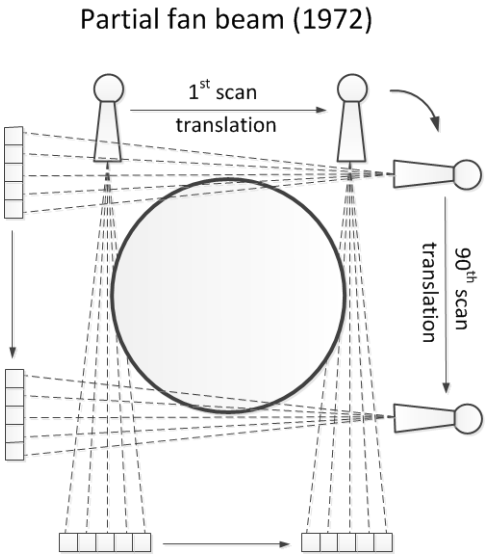
which used divergent-beam/fan-beam geometry. It used the source translate-rotate technology as the first generation but had multiple detectors and a fan-shaped beam. The later two generations of scan configurations led to even faster data collection. For the third generation, both the source and the detector array were mounted on a yoke which rotated around the patient over 360°. Data collection time for such scanners ranged from 1 to 20 seconds. For the fourth generation, a large number of detectors are mounted on a fixed ring. Inside this ring there is an X-ray tube continually rotating around the patient. During this rotation the output of the detector integrators facing the tube is sampled every few milliseconds. In 1989, the first investigations and clinical trials of spiral/helical CT were already completed, which symbolizes the transition from slice-by-slice imaging to true volume imaging. The new machines are equipped with a very large number of detector rows and with scanners based on cone-beam geometry, as shown in Figure 2.3. The grid rotates around an axis in the gantry. The table and the patient are placed in the gantry along this axis of rotation. Depending on the motion of the table along the axis, helical or axial scans can be obtained. Fully 3-D image reconstruction becomes possible as it permits continuous rotation with the object being imaged slowly and smoothly slid through the X-ray ring. Hence, modern multi-slice CT scanners allow the performance of most desired examinations with very high reliability.

According to Kalender [78], "CT has shown a steady upward trend with respect

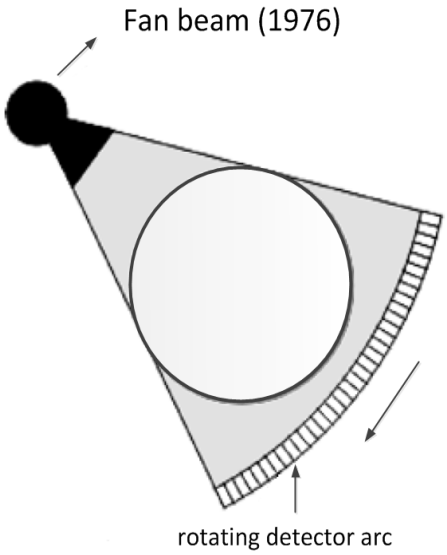
to technology, performance and clinical use. The upward trend is unbroken for the time being, and the position of CT in clinical radiology appears consolidated to a higher degree than ever before."



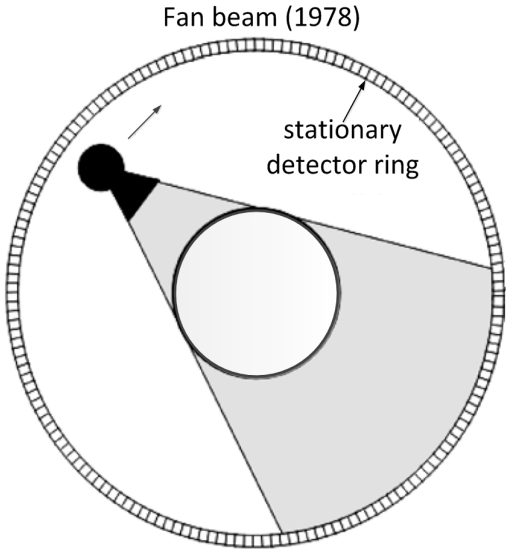
(a) *First generation: translation/rotation*



(b) *Second generation: translation/rotation*



(c) *Third generation: continuous rotation*



(d) *Fourth generation: continuous rotation*

Figure 2.2: Four scanner generations since the 1970s

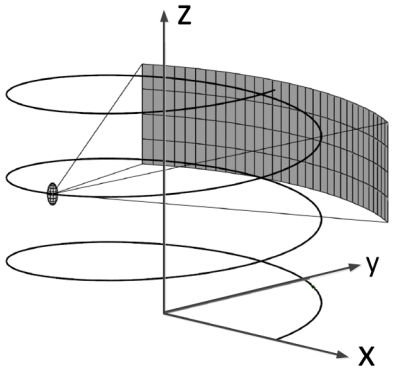


Figure 2.3: Cone-beam acquisition geometry

2.2 Physics

In the CT scanning shown in Figure 2.4, the beam attenuates due to the photoelectric and Compton effects, respectively. Photoelectric absorption involves an X-ray photon imparting all its energy to a tightly bound inner electron in an atom. The electron uses some of this acquired energy to overcome the binding energy within its shell, the rest appearing as the kinetic energy of the thus freed electron. On the other hand, Compton scattering involves the interaction of the X-ray photon with either a free electron, or one that is only loosely bound in one of the outer shells of an atom. Because of this interaction, the X-ray photon is deflected from its original direction of travel with some loss of energy, which is gained by the electron [77].

Assume that the X-ray beam only consists of monoenergetic photons, let N_{in} be the total number of photons entering the object, and N_{out} be the total number of photons exiting through the beam on the detector side. When the beam width is sufficiently small, the relationship between N_{in} and N_{out} is revealed in [60]:

$$N_{out} = N_{in} e^{-\int_{ray} \mu(x,y) dr}$$

where $\mu(x, y)$ denotes the attenuation function of the object with two space coordinates x and y, and r is an element of length along the ray.

In practice, the X-ray source used for medical imaging is polychromatic rather

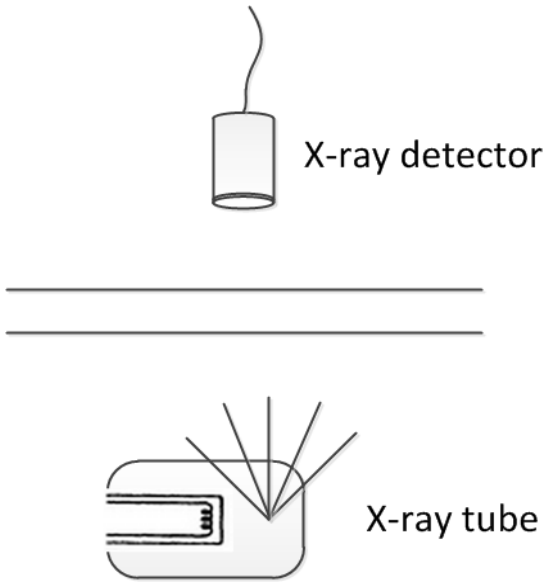


Figure 2.4: CT scanning: the beam is measured on the detector side to determine the attenuation of the object.

than monochromatic. The above equation no longer holds and is replaced by:

$$S_{out}(E) = S_{in}(E)e^{-\int_{ray}\mu(x,y,E)dr}$$

where E stands for the energy, $S_{in}(E)$ represents the incident photon number density, also known as the energy spectral density of the incident photons, and $S_{out}(E)$ is obviously the density of the exiting photons.

CT images are maps of the relative linear attenuation values of tissues. The outputs of the computer attached to the scanner are integers called CT numbers, also known as Hounsfield units (HUs). A HU is a function of the material attenuation coefficient, defined as:

$$HU = \frac{\mu - \mu_{water}}{\mu_{water}} \times 1000$$

where μ is the attenuation coefficient of the material and μ_{water} is that of water.

Various artifacts are commonly encountered in clinical CT and we briefly review the cause and appearance of each type [16].

Ring artifact: A mis-calibrated or defective detector element results in rings centred on the centre of rotation, hence the name. These can often be fixed by recalibrating the detector.

Noise: The statistical error of low photon counts causes Poisson noise, which results in random thin bright and dark streaks that appear preferentially along the direction of the greatest attenuation. It can be reduced using noise reduction techniques, or iterative reconstruction, or by combining data from multiple scans.

Beam hardening and scatter: Beam hardening is caused by the fact that the linear attenuation coefficient for many tissues decreases with energy in the energy ranges used for diagnostic examinations. The low energy photons are preferentially absorbed, making the mean energy associated with the exit spectrum higher than that associated with the incident spectrum, i.e. the beam-hardening phenomenon. Both beam hardening and scatter cause pseudo-enhancement of renal cysts and produce dark streaks between two high attenuation objects, e.g. metal and bone, with surrounding bright streaks. Iterative reconstruction can help to reduce them.

Metal artifact: Multiple mechanisms can cause metal streak artifacts, including beam hardening, scatter, Poisson noise, motion, and edge effects. The metal deletion technique is an iterative method that reduces artifacts and is quite influential on the diagnosis.

2.3 Low-radiation CT

Despite the overwhelming benefits of CT for patients, concerns have been raised regarding the potential risk of cancer from CT because the increased X-ray radiation can lead to ionization of body cells. Since its introduction in 1973, CT has established itself as a primary diagnostic imaging modality due to its tremendous value. It is estimated that 67 million CT examinations were performed in 2006 in the USA [1], up from 3 million in 1980 [18]. Because of its pervasive use in modern medical practice, CT alone contributed almost one half of the total radiation exposure from medical use and one quarter of the average radiation exposure per capita in 2009 in the USA [1]. Hence, it is clear that reducing radiation dose ranks one of the top priorities of the CT community. Two guiding principles must be followed [103]. First, CT examinations must be fully justified for each individual patient. Second, for each CT examination, all technical aspects of the examination must be optimized, such that the required level of image quality can

be obtained while keeping the doses as low as possible [164].

To quantify CT radiation dose, several dose metrics are adopted: scanner radiation output, organ dose and effective dose. Note that these metrics are not a direct measurement of patient dose, but rather a standardized dose metric of scanner output levels applying to standardized phantoms. To be specific, the scanner radiation output is measured based on two acrylic phantoms: head and body, currently represented by the volume CT dose index ($CTDI_{vol}$). Organ dose quantifies the radiation risk to a specific organ, and effective dose, typically expressed in the units of mSv, represents the 'whole-body equivalent' dose that would have a similar risk of health detriment as that due to partial body irradiation. There are considerable uncertainties in evaluating the risk of developing a radiation-induced cancer when the dose is lower than 100 mSv. Since the radiation dose is closely related to the CT image quality, keeping it as low as reasonably achievable and consistent with the diagnostic task remains the most important strategy for decreasing this potential risk [164]. Image quality can be evaluated via several metrics: noise, high-contrast and low-contrast spatial resolution. Therefore, it is crucial to recognize the requirements of specific diagnostic tasks in approaching the goal of reducing radiation.

Various dose-management strategies for different types of CT have been proposed and are reviewed below:

- CT system optimization

  The dose efficiency of a CT system mainly depends on the detector, collimator and beam-shaping filter. Two dose-relevant characteristics of a detector are its quantum detection efficiency and geometrical efficiency, which together describe the effectiveness of the detector in converting incident X-ray energy into signals. The design of the collimator can define the X-ray beam accordingly and avoid unnecessary radiation exposure. Recent progress involves reducing the amount of over-scanning in spiral CT, with up to 40% of dose reduction reported [34]. The X-ray beam filter is a physical object that attenuates and 'hardens' the beam spectra to guarantee that the beam is hard enough to efficiently penetrate the patient and give sufficient contrast information. In addition, the current increasing use of dual-energy systems for CT is also helpful for soft tissue segmentation in medical imaging.

- Scan range

  The scan range should be chosen carefully to cover the region of interest and avoid the direct radiation exposure of other regions. It is a quite straightforward way to keep the radiation low.

- Automatic exposure control

  Automatic exposure control (AEC) aims to automatically modulate the tube current to accommodate differences of patient anatomy, shape and size, taking the projection angle, longitudinal location along the patient or both as variables. This near-real time system can be fully preprogrammed and has been adopted by the major manufacturers.

- Optimal tube potential

  Recent physics and clinical studies [66, 150] have demonstrated that using lower tube potentials (kV) can achieve dose reduction and contrast enhancement, but will result in noisier images. Therefore, there is a trade-off between image noise and contrast in determining the tube potential value, which is highly dependent on patient size and the diagnostic requirements.

- Data processing

  Dose efficiency can be improved via data processing techniques: denoise, reconstruction and modelling. To control noise in CT, the filters are applied to the raw/log-transformed sinogram or the reconstructed images. For reconstruction, iterative algorithms are adopted instead of analytic algorithms, since the former can recover the target from much fewer projections. Finally, in order to determine the lowest possible radiation dose, researchers simulate CT examinations at different dose levels by inserting realistic quantum and electronic noise and decide the dose according to the diagnostic task.

In addition to the general strategies described above, each CT clinical application has its own considerations and limitations for dose reduction. The future perspective of radiation dose reduction includes individualizing scanning techniques, novel reconstruction methods, photon-counting detectors, etc. To conclude, the guiding principle for a medical CT examination is to keep the radiation dose as low as reasonably achievable.

2.4 Analytical Reconstruction

The mathematical foundation of tomographic reconstruction dates back to 1917 when an Austrian mathematician Johann Radon invented the Radon transform [122], which states that an object can be reconstructed from an infinite set of its projections and relates the projections and the object in the frequency domain. In 1937, a Polish mathematician, Stefan Kaczmarz, developed a method to find an approximate solution to a large system of linear algebraic equations [75], which laid down the basis for the powerful iterative reconstruction method. It was later adapted by Sir Godfrey Hounsfield as the image reconstruction mechanism in the first commercial CT scanner. Despite its early success, it was later replaced by analytical methods due to its very high computational demands when the amount of measured data increased, but has had a renaissance in recent years. The analytical and iterative reconstruction algorithms are reviewed in Section 2.4 and Section 2.5, respectively.

2.4.1 Fourier Slice Theorem

We start with the mathematical basis of tomography and give the definition of line integrals that compose the projections. The coordinate system is defined in Figure 2.5, where the object is represented by a 2-D function $f(x, y)$ and each line integral by the angular and radial coordinates (θ, r).

The line integral $s(\theta, r)$ can be defined as

$$s(\theta, r) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\delta(r - x\cos\theta - y\sin\theta)dxdy \tag{2.1}$$

where $\delta(\cdot)$ is the Dirac delta function. The projection $s(\theta, r)$ is known as the Radon transform of the object function $f(x, y)$.

Conventional X-ray CT image reconstruction relies on analytical image reconstruction methods, among which filtered back-projection (FBP) is the most popular. FBP is based on the Fourier slice theorem (FST), which proves that the 1-D Fourier transform of the projections is equal to the 2-D Fourier transform of the image evaluated on the line that the projection was taken. The derivation details of FST are revealed below. The 1-D Fourier transform of the Radon transform

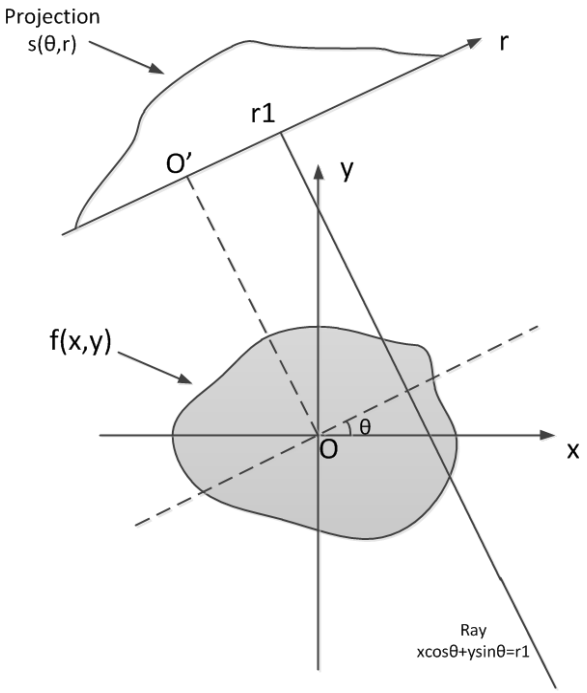


Figure 2.5: The coordinate system

with regard to r is given by:

$$\hat{S}(\theta, \rho) = \int_{-\infty}^{\infty} s(\theta, r) e^{-jr\rho} dr$$

with ρ as the index in frequency domain. Substituting (2.1) into the above equation we find:

$$\hat{S}(\theta, \rho) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) e^{-j\rho(x\cos\theta + y\sin\theta)} dxdy \tag{2.2}$$

Compared with the 2-D Fourier transform of the object $f(x, y)$:

$$\hat{F}(u, v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) e^{-j(ux+vy)} dxdy \tag{2.3}$$

with u, v as the index in frequency domain, it is easy to conclude that:

$$\hat{S}(\theta, \rho) = \hat{F}(\rho\cos\theta, \rho\sin\theta) \tag{2.4}$$

This leads to the Fourier slice theorem stated below:

Theorem 2.1. *[77] The Fourier transform of a parallel projection of an image $f(x, y)$ taken at angle θ gives a slice of the two-dimensional transform $\hat{F}(u, v)$,*

subtending an angle θ with the u-axis.

For illustration, it is shown in Figure 2.6 that the Fourier transform of $s(\theta, r)$ gives the values of $\hat{F}(u, v)$ along line AB.

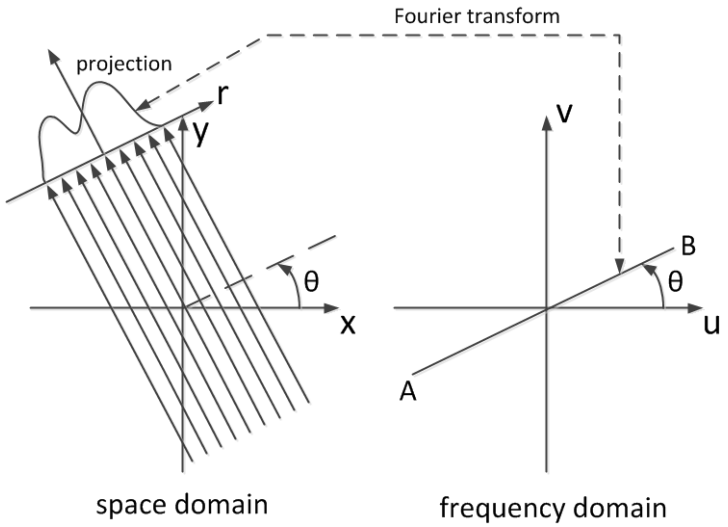


Figure 2.6: Illustration of FST

2.4.2 FBP for Parallel-beam and Fan-beam CT

In this section, we simply give the resulting analytical inversion formula, as its derivation is quite straightforward and interested readers are referred to [77]. Consider the parallel-beam CT in Figure 2.7. Suppose that $\hat{S}(\theta, \rho)$ is the 1-D Fourier transform of a projection $s(\theta, r)$ with respect to r, we have:

$$f(x, y) = \int_0^{\pi} [\int_{-\infty}^{\infty} \hat{S}(\theta, \rho)|\rho| e^{j\rho r} d\rho] d\theta, \quad r = x\cos\theta + y\sin\theta$$

Assuming there are q views of projections, the implementation of the FBP algorithm is presented below:

FBP algorithm
measure the projection $s(\theta, r)$
apply 1-D Fourier transform to the projection w.r.t. r and obtain $\hat{S}(\theta, \rho)$
multiply $\hat{S}(\theta, \rho)$ with the weighting function $2\pi\|\rho\|/q$
sum over the image plane the inverse Fourier transform of the filtered projections

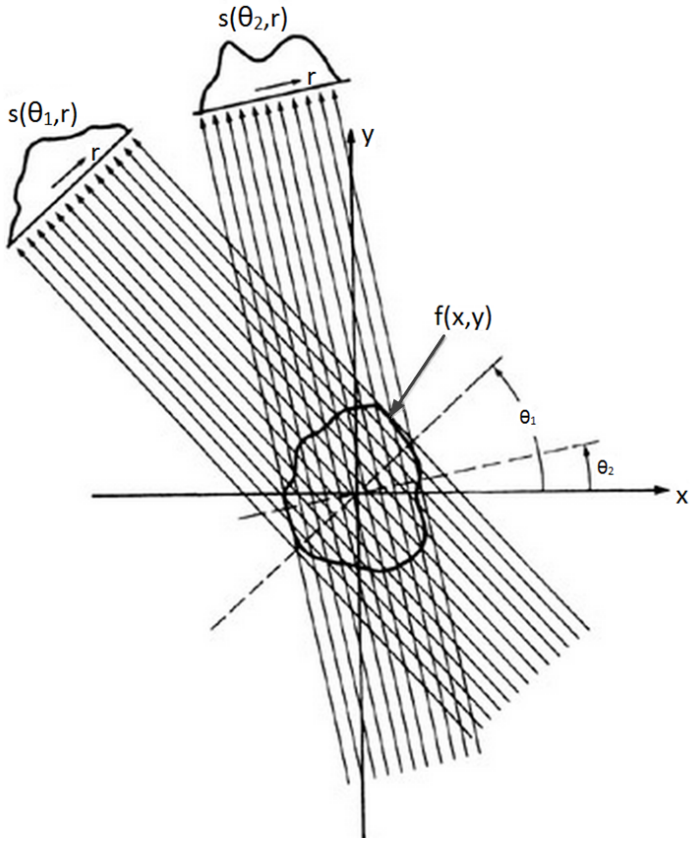


Figure 2.7: Parallel projections

Consider the fan-beam CT with equiangular rays illustrated in Figure 2.8. The source travels along a circular trajectory, with R as the distance from the object centre and angle β to define its position. The curved detector array lies on a circle of radius D centred at the source. Any ray in the fan beam is specified by an angle $\gamma \in [-\gamma_m, \gamma_m]$ that increases in the clockwise direction. Then the standard FBP formula is [38]:

$$f(x,y) = \frac{1}{2}\int_0^{2\pi} d\beta \frac{R}{||(x-R\cos\beta, y-R\sin\beta)^T||^2} \int_{-\gamma_m}^{\gamma_m} d\gamma h_{ramp}(\sin(\gamma^* - \gamma)) \cos\gamma s_f(\beta,\gamma)$$

where h_{ramp} is the ramp filter, and γ^* is the value of γ for the ray that connects the X-ray source with $(x,y)^T$:

$$\gamma^* = \tan^{-1}(u^*/D) \quad with \quad u^* = \frac{D(-x\sin\beta + y\cos\beta)}{R - x\cos\beta - y\sin\beta}$$

The reconstruction speed of FBP is quite fast due to its simplicity and image quality is usually satisfactory. However, a high volume of projections is required by Shannon-Nyquist theorem [136]. As stated in [77], for a well-balanced $n \times n$

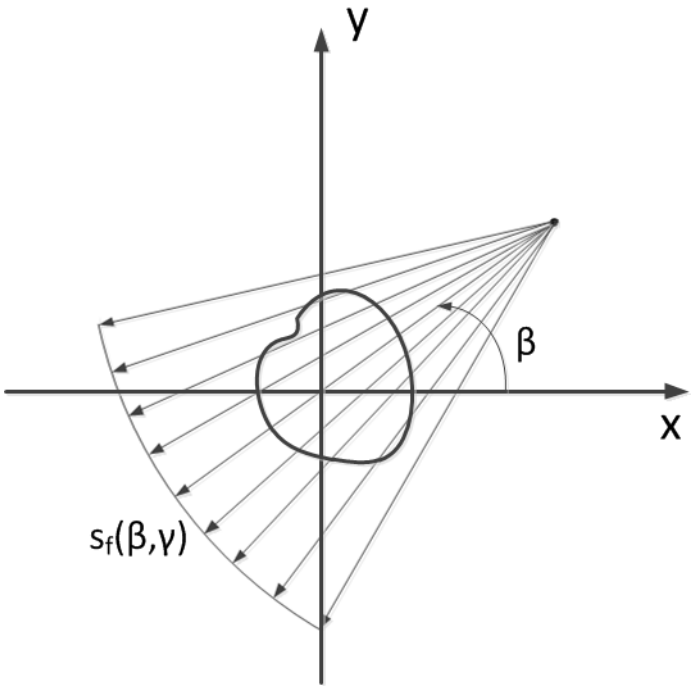


Figure 2.8: Fan-beam projections

reconstructed image, the number of rays in each projection should be roughly n and the number of views should also be roughly n.

2.4.3 FDK and Katsevich Formula for Cone-beam CT

The interest in fully 3-D CT reconstruction increased with the development of computing power. In 1984, Feldkamp, Davis, and Kress [49] proposed a filtered back-projection algorithm for 3-D axial cone-beam CT, which was known as FDK and later extended to more general scanning paths, such as the helix [151]. Throughout the present thesis, bold case letters/symbols denote vectors. The local detector coordinate is defined in Figure 2.9, where the source curve is a helix, given by:

$$\boldsymbol{y}(\phi) = [R\cos\phi \quad R\sin\phi \quad P\frac{\phi}{2\pi}]^T, \quad \phi \in [0 \quad 2\pi)$$

where R is the helical radius and P is the helical pitch, i.e. the displacement of the patient table per source turn. For a flat detector,

$$\mathbf{e}_b(\phi) = [-\sin\phi \quad \cos\phi \quad 0]^T$$

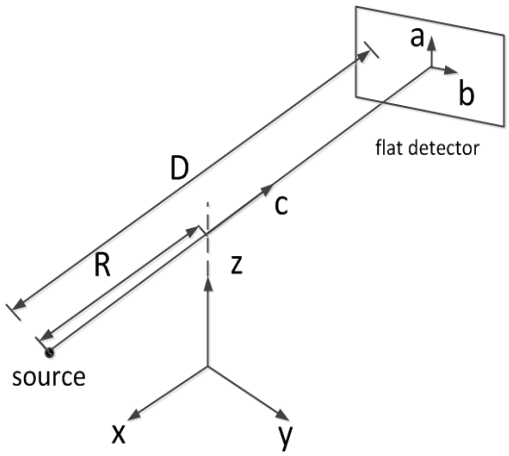


Figure 2.9: Flat detector geometry

$$\mathbf{e}_c(\phi) = [-\cos\phi \quad -\sin\phi \quad 0]^T$$

$$\mathbf{e}_a = [0 \quad 0 \quad 1]^T$$

where $\mathbf{e}_b(\phi)$ and $\mathbf{e}_a$ span the detector, and $\mathbf{e}_c(\phi)$ points from the source to the detector centre. (b^*, a^*) are the coordinates for the point of intersection of the detector with the line passing through the source and $\mathbf{e} = [x_0, y_0, z_0]^T$.

$$\begin{aligned} b^* &= \frac{D}{c^*(\phi, \mathbf{e})}(-x_0 \sin\phi + y_0 \cos\phi) \\ a^* &= \frac{D}{c^*(\phi, \mathbf{e})}(z_0 - \frac{P\phi}{2\pi}) \end{aligned}$$

where $c^*(\phi, \mathbf{e}) = R + < \mathbf{e}, \mathbf{e}_c >= R - x_0 \cos\phi - y_0 \sin\phi$. The unit vector which defines the direction from the source to the detector at (b, a) is:

$$\begin{aligned} \boldsymbol{\theta}(\phi, b, a) = & \quad \frac{1}{\sqrt{b^2 + D^2 + a^2}}(b\mathbf{e}_b(\phi) + D\mathbf{e}_c(\phi) + a\mathbf{e}_a) \\ = \frac{1}{\sqrt{b^2+D^2+a^2}} & \quad [-b\sin\phi - D\cos\phi \quad b\cos\phi - D\sin\phi \quad a]^T \end{aligned}$$

Given (ϕ, b, a), the X-ray projection data of the object $f(\boldsymbol{e})$ with these local coordinates can be expressed as:

$$g_f(\phi, b, a) = Df(\boldsymbol{y}, \boldsymbol{\theta}) = \int_0^\infty f(\boldsymbol{y} + t\boldsymbol{\theta})dt \tag{2.5}$$

where Df is the cone-beam transformation defined in $\mathbb{R}^3$.

Then the FDK method for a flat detector with a helical scanning path is formu-

lated as:

$$f(\mathbf{e}) = \int_0^{2\pi} \frac{D^2}{(c^*)^2} \int_{-b_0}^{b_0} k_b(b^* - b') g_f(\phi, b', a^*) \frac{D db' d\phi}{\sqrt{D^2 + (b')^2 + (a^*)^2}},$$

where k_b is a filtering kernel (e.g. Shepp-Logan) appropriate for 2-D reconstruction, and $g_f(\phi, b, a)$ is the cone-beam data in local coordinates. The role played by the filtering kernel is to reduce the blurring effect caused by the back-projection. The implementation of FDK is straightforward and similar to that of FBP.

The FDK algorithm is fast and gives reasonable results in some circumstances, but it is difficult to fully analyse its numerical properties and to predict reconstruction artifacts, as the algorithm is based on an approximate inversion formula. Since the FDK algorithm was proposed, a great deal of effort has been expended in investigating faster and more accurate reconstruction methods. Alexander Katsevich achieved a breakthrough in 2002 with a series of papers [79–82], in which he proved a theoretically exact reconstruction formula of the filtered back-projection type for helical cone-beam CT. The Katsevich formula makes use of the π-line and κ-plane, where the π-line is defined to be any line segment that connects two points on the helix which are separated by less than one helical turn, and the κ-plane is any plane that has three intersections with the helix such that one intersection is half-way between the two others. Let $\boldsymbol{y}(\phi_1)$ and $\boldsymbol{y}(\phi_2)$ be the endpoints of the π-line lying on the helix, then $I_\pi = [\phi_1, \phi_2]$ is the parametric interval corresponding to the unique π-line passing through $\mathbf{e}$. $\boldsymbol{m}(\phi, \boldsymbol{\beta})$ is defined to be a unit normal vector for the plane $\kappa(\phi, \psi)$ with the smallest $|\psi|$ value that contains the line of direction $\boldsymbol{\beta}$, where $\boldsymbol{\beta}(\phi, \boldsymbol{e}) = \frac{\boldsymbol{e} - \boldsymbol{y}(\phi)}{||\boldsymbol{x} - \boldsymbol{y}(\phi)||}$, as the unit vector pointing from $\boldsymbol{y}(\phi)$ to $\boldsymbol{e}$. Hence, the Katsevich formula can be stated as:

$$f(\boldsymbol{e}) = -\frac{1}{2\pi^2} \int_{I_\pi(\boldsymbol{e})} \frac{1}{||\boldsymbol{e} - \boldsymbol{y}(\phi)||} \mathrm{PV} \int_0^{2\pi} \frac{\partial}{\partial q} D f(\boldsymbol{y}(q), \boldsymbol{\alpha}(\phi, \boldsymbol{e}, \gamma))|_{q=\phi} \frac{d\gamma d\phi}{\sin\gamma} \quad (2.6)$$

where PV means that the integral should be interpreted as a principal value integral, $\boldsymbol{\alpha}(\phi, \boldsymbol{e}, \gamma) = \cos\gamma \boldsymbol{\beta}(\phi, \boldsymbol{e}) + \sin\gamma(\boldsymbol{\beta}(\phi, \boldsymbol{e}) \times \boldsymbol{m}(\phi, \boldsymbol{\beta}))$. (2.6) is the general reconstruction formula disregarding the detector geometry, and Noo et al. [112] implemented Katsevich's formula of direct reconstruction from data measured in flat detector coordinates, shown as below:

$$f(\boldsymbol{e}) = \frac{1}{2\pi} \int_{I_\pi(\boldsymbol{e})} \frac{1}{c^*(\phi, \boldsymbol{e})} \boldsymbol{P^F}(\phi, b^*, a^*) d\phi \quad (2.7)$$

where

$$\boldsymbol{P^F}(\phi, b^*, a^*) = \int_{-\infty}^{\infty} \frac{1}{\pi(b^* - (b^*)')} \frac{D}{\sqrt{(b^*)'^2 + D^2 + (a^*)'^2}} \boldsymbol{P_1}(\phi, (b^*)', (a^*)') d(b^*)'$$

$$\begin{aligned} \boldsymbol{P_1}(\phi, (b^*)', (a^*)') &= \left(\frac{\partial \boldsymbol{P}(q, (b^*)', (a^*)')}{\partial q}\right. \\ &+ \frac{((b^*)')^2 + D^2}{D} \frac{\partial \boldsymbol{P}(q, (b^*)', (a^*)')}{\partial (b^*)'} \\ &+ \left.\frac{(b^*)'(a^*)'}{D} \frac{\partial \boldsymbol{P}(q, (b^*)', (a^*)')}{\partial (a^*)'}\right)\Big|_{q=\phi} \end{aligned}$$

The detailed implementation of this algorithm can be found in [155].

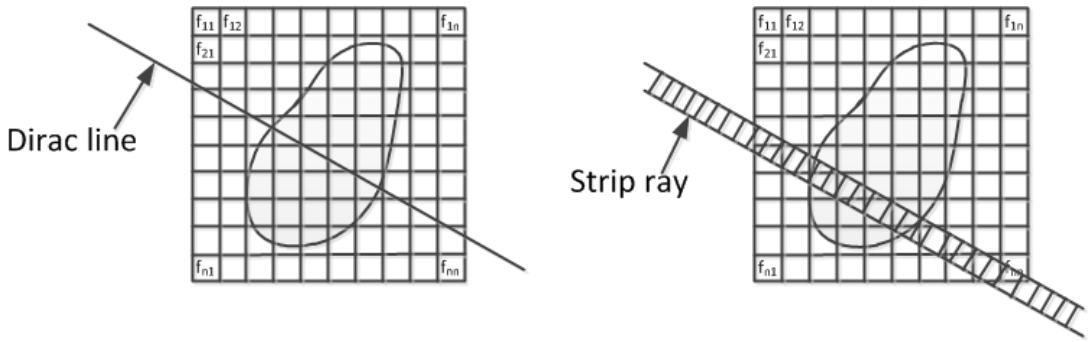
The methods mentioned above are quite mature nowadays, and it is possible to reconstruct an image volume with typical image size in a minute with modern computer architecture.

2.5 Iterative Reconstruction

2.5.1 Projection Models

The forward projection operator or the CT system matrix, A, defined to transform the object to a set of line integrals, is the key element in deciding reconstruction accuracy via iterative methods. It defines (1) how the continuous function of the object is represented by a finite set of parameters; and (2) how projection data are calculated from this continuous function. Correspondingly, forward projection models vary on the choice of the image basis function that describes the pixel/voxel shapes and the integration function that is related to the acquisition geometry [97]. Many basis functions have been proposed, such as square basis function [118, 145], Fourier series, circular harmonics, wavelets, of which the square basis function is the most popular and common for its simplicity. Another one that has gained favourable results in CT reconstruction is the Kaiser-Bessel function, also known as 'blob'. It assumes that non-zero values only exist in a circular disk around the origin, and smoothly decrease from a positive value at the origin to zero at the edge of the disk. 'blob' has been a particularly popular

choice of rotationally symmetric basis. For the integration functions, there are also several types, among which the Dirac line is the most common. Its geometry is shown in Figure 2.10(a), and the sinogram data is gained by the integral along the line. Other types of integration functions that consist either of several Dirac lines or a strip (Figure 2.10(b)), can be used for suppressing aliasing in the single Dirac line type. Hence, a variety of forward projection models can be defined as a combination of a selected basis function and an appropriate geometrical integration function. Table 2.1 lists the most popular forward projection methods. The linear intersection model, which is a variant of ray-driven methods and is used in the simulations in the present thesis, is described in detail below.



(a) *Projection model with Dirac line* (b) *Projection model with strip ray*

Figure 2.10: Projection model with different integration functions. In model (a), a popular choice of the weight for each pixel is the intersection length. In model (b), the ray width is usually equal to the pixel width and a popular choice of the weight for each pixel is the intersection area.

In the CT system, the line integral is known as the Radon transform of the object f, denoted by:

$$s(\theta, r) = \int_x \int_y f(x, y)\delta(x\cos\theta + y\sin\theta - r)dxdy$$

where (θ, r) is the coordinate of parallel-beam CT, defined in Figure 2.5. As the Radon transform is continuous both for the object under scan and the projections themselves, it has to be sampled to adjust to practical applications. Suppose the object is discretized into $n \times n$ small lattices, and the gray-scale value keeps the

Table 2.1: Selection of the most prominent projection methods and references to initial or important publications

Projection model	Method	Reference
Ray-driven methods	Trace the ray path at each projection bin.	[64, 137, 165, 167]
Pixel-driven methods	For each image pixel, the centre of the pixel is projected onto the detector array along the projection direction, and a value is obtained from, or accumulated in, the detector by (typically linear) interpolation.	[64, 118, 167]
Distance-driven methods	Map pixel and detector boundaries to a common axis and the coefficients are computed as the row or slab intersection length combined with the overlap coefficient (the length of overlap).	[99, 100]

same within each lattice. The matrix form of the object is:

$$f = \begin{pmatrix} f_{11} & f_{12} & \cdots & f_{1n} \\ f_{21} & f_{22} & \cdots & f_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ f_{n1} & f_{n2} & \cdots & f_{nn} \end{pmatrix}$$

The projection becomes a so-called ray-sum:

$$s(\theta, r) = \sum_{i=1}^{n} \sum_{j=1}^{n} f_{ij} l_{ij}$$

where $l_{ij} = \int_x \int_y \delta(xcos(\theta) + ysin(\theta) - r)dxdy$. In the linear intersection model, as the name implies, it is the intersection length of the X-ray and the lattice at the ith row and jth column.

For the application of compressed sensing, the reconstructed problem is formulated as:

$$\mathbf{b} = A\mathbf{x} \tag{2.8}$$

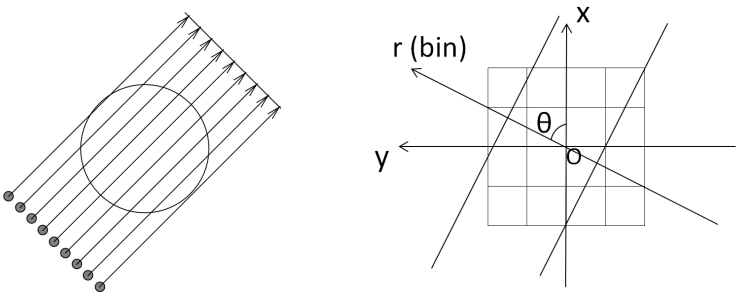
where $\mathbf{x}$ is the column-wise vector of the object f,

$$\mathbf{x} = [f_{11} \quad f_{21} \cdots f_{n1} \quad f_{12} \quad f_{22} \cdots f_{n2} \cdots f_{1n} \quad f_{2n} \cdots f_{nn}]^T$$

$\mathbf{b}$ is the projection vector with the size $(q \times p) \times 1$, q is the number of the scanning views and p stands for the number of rays at each view.

$$\mathbf{b} = [\underbrace{b_{11} \quad b_{12} \cdots b_{1p}}_{\theta_1} \quad \underbrace{b_{21} \quad b_{22} \cdots b_{2p}}_{\theta_2} \quad \cdots \quad \underbrace{b_{q1} \quad b_{q2} \cdots b_{qp}}_{\theta_q}]^T$$

Obviously, A is of the size $(q \times p) \times n^2$, which is extremely large. We start the formulation of A by building the discretized model of parallel-beam CT as in Figure 2.11.



(a) *Parallel-beam scanning*

(b) *Coordinates*

Figure 2.11: Discretized model of parallel-beam CT

$A((i-1) \times p + j, l)$ denotes the length of the jth ray at the ith angle through the lth pixel of $\mathbf{x}$. The lth pixel in $\mathbf{x}$ corresponds to 4 points in the model of Figure 2.11(b): $(x_1, y_1), (x_1, y_2), (x_1, y_1), (x_2, y_1)$, where

$$x_1 = \frac{n}{2} - e \quad x_2 = \frac{n}{2} - e + 1 \quad y_1 = \frac{n}{2} - c \quad y_2 = \frac{n}{2} - c + 1$$

$$e = mod(l, n), c = \frac{l - e}{n} + 1.$$

Now we can calculate the y-coordinates of the intersections of the ray with two lines $x = x_1, x = x_2$ as

$$\begin{cases} y_1' & = r(j) \cdot \sin\theta_i - \frac{(x_1 - r(j) \cdot \cos\theta_i)}{\tan\theta_i} \\ y_2' & = r(j) \cdot \sin\theta_i - \frac{(x_2 - r(j) \cdot \cos\theta_i)}{\tan\theta_i} \end{cases}$$

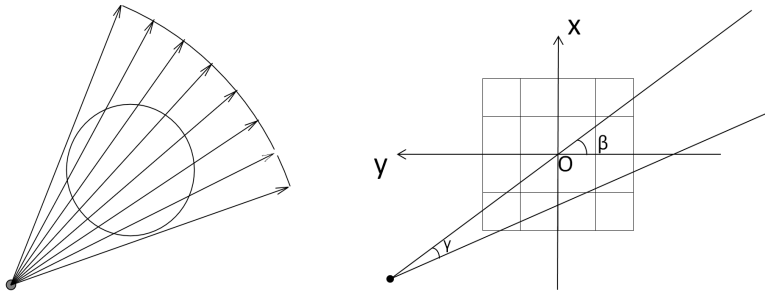
and the x-coordinates of the intersections of the ray with two lines $y = y_1, y = y_2$ as

$$\begin{cases} x_1' &= r(j) \cdot \cos\theta_i - \tan\theta_i (y_1 - r(j) \cdot \sin\theta_i) \\ x_2' &= r(j) \cdot \cos\theta_i - \tan\theta_i (y_2 - r(j) \cdot \sin\theta_i) \end{cases}$$

Here $r(j)$ is the the bin coordinate of the jth ray. Hence,

$$A((i-1) \times p + j, l) = \begin{cases} \sqrt{(x_1 - x_1')^2 + (y_1 - y_1')^2} & \text{if} \quad x_1 \le x_1' \le x_2 \quad \& \quad y_1 \le y_1' \le y_2 \\ \sqrt{(x_2 - x_1')^2 + (y_1 - y_2')^2} & \text{if} \quad x_1 \le x_1' \le x_2 \quad \& \quad y_1 \le y_2' \le y_2 \\ \sqrt{(x_1 - x_2')^2 + (y_2 - y_1')^2} & \text{if} \quad x_1 \le x_2' \le x_2 \quad \& \quad y_1 \le y_1' \le y_2 \\ \sqrt{(x_2 - x_2')^2 + (y_2 - y_2')^2} & \text{if} \quad x_1 \le x_2' \le x_2 \quad \& \quad y_1 \le y_2' \le y_2 \\ \sqrt{(x_1' - x_2')^2 + 1} & \text{if} \quad x_1 \le x_1' \le x_2 \quad \& \quad x_1 \le x_2' \le x_2 \\ \sqrt{(y_1' - y_2')^2 + 1} & \text{if} \quad y_1 \le y_1' \le y_2 \quad \& \quad y_1 \le y_2' \le y_2 \\ 0 & \text{else} \end{cases}$$

The discretized model of fan-beam CT is shown in Figure 2.12. Suppose that the number of rays at each view is p, and the distance from the source to the centre of the domain O is R.



(a) *Fan-beam scanning* (b) *Coordinates*

Figure 2.12: Discretized model of fan-beam CT

When the source is at the ith angle, the y-coordinates of the intersections of the jth ray with two lines $x = x_1, x = x_2$ are

$$\begin{cases} y_1' &= R\cos\beta_i - \frac{x_1 + R\sin\beta_i}{\tan(\beta_i + \gamma_j)} \\ y_2' &= R\cos\beta_i - \frac{x_2 + R\sin\beta_i}{\tan(\beta_i + \gamma_j)} \end{cases}$$

and the x-coordinates of the intersections of the jth ray with two lines $y = y_1, y =$

y_2 are

$$\begin{cases} x_1' &= (R\cos\beta_i - y_1)\cdot\tan(\beta_i+\gamma_j) - R\sin\beta_i \\ x_2' &= (R\cos\beta_i - y_2)\cdot\tan(\beta_i+\gamma_j) - R\sin\beta_i \end{cases}$$

Hence

$$A((i-1)\times p+j,l) = \begin{cases} \sqrt{(x_1-x_1')^2+(y_1-y_1')^2} & \text{if} \quad x_1 \le x_1' \le x_2 \quad \& \quad y_1 \le y_1' \le y_2 \\ \sqrt{(x_2-x_1')^2+(y_1-y_2')^2} & \text{if} \quad x_1 \le x_1' \le x_2 \quad \& \quad y_1 \le y_2' \le y_2 \\ \sqrt{(x_1-x_2')^2+(y_2-y_1')^2} & \text{if} \quad x_1 \le x_2' \le x_2 \quad \& \quad y_1 \le y_1' \le y_2 \\ \sqrt{(x_2-x_2')^2+(y_2-y_2')^2} & \text{if} \quad x_1 \le x_2' \le x_2 \quad \& \quad y_1 \le y_2' \le y_2 \\ \sqrt{(x_1'-x_2')^2+1} & \text{if} \quad x_1 \le x_1' \le x_2 \quad \& \quad x_1 \le x_2' \le x_2 \\ \sqrt{(y_1'-y_2')^2+1} & \text{if} \quad y_1 \le y_1' \le y_2 \quad \& \quad y_1 \le y_2' \le y_2 \\ 0 & \text{else} \end{cases}$$

This concludes the formulation of the linear intersection model for parallel-beam and fan-beam CT, and it will be used in the simulations in Chapter 5.

2.5.2 Algorithms

Iterative reconstruction methods are entirely different from analytical methods and conceptually easier, and involve assuming the object contains an array of unknowns and then setting up algebraic equations for the unknowns in terms of the measured projection data. The availability of large computational capacities in normal workstations and the ongoing efforts towards lower doses in CT have made the iterative reconstruction a hot topic for all major vendors of clinical CT systems in the past few years. A brief overview of iterative reconstruction algorithms is presented below.

Suppose A is the CT system matrix, $\mathbf{x}$ is the object vector to be reconstructed, and $\mathbf{b}$ is the measured data, one obtains:

$$A\mathbf{x} = \mathbf{b} \tag{2.9}$$

The algorithms developed to solve (2.9) fall into the first category of iterative image reconstruction algorithms, referred to as the algebraic reconstruction technique (ART) [57]. Different constraints can be imposed upon the iterations, such as positivity constraint, prior information, etc. While the original ART

works on single rays and thus single pixels, there have been efforts to develop its variants, aiming at higher efficiency. For example, the simultaneous algebraic reconstruction technique (SART) [5, 71] is used to perform updates for complete projections, the ordered subsets (OS) method [101] divides the projections into subsets and updates each group, and the multiplicative algebraic reconstruction technique (MART) [7] multiplies the update term onto the current solution instead of adding or subtracting.

The entire process of data acquisition described in (2.9) is modeled as a deterministic process, which cannot account for the intrinsic photon statistics that introduce noise into the measurements [143]. Hence, the other category of iterative image reconstruction algorithms, called the statistical method, utilizes knowledge of the underlying physics, i.e. the understanding of the statistical distribution resulting from the X-ray interaction process. The statistical method is generally formulated as:

$$\min_{\mathbf{x}} \quad \frac{1}{2}(\mathbf{b} - A\mathbf{x})^T Q(\mathbf{b} - A\mathbf{x}), \tag{2.10}$$

where Q is the diagonal matrix with ith element of σ_i^2, i.e. an estimate of the variance of noise of line integral at detector bin i. Here σ_i^2 is calculated using

$$\sigma_i^2 = \frac{e^{\bar{b}_i}}{N_{i0}},$$

where $\bar{b}_i$ is the mean of the noisy sinogram datum b_i, and N_{i0} is the incident photon number at detector bin i. This category can be roughly divided into two groups of methods that are based on the maximum likelihood (ML) principle and the least squares (LS) principle, respectively [10]. The maximum likelihood expectation-maximization (ML-EM) algorithm consists of two alternating steps that compute the expectation of the log-likelihood and find the next estimate by maximizing the expected log-likelihood. Again, the OS principle is applied for faster convergence. For the least squares principle, methods based on iterative coordinate descent (ICD) [133, 145] have been proposed and successfully implemented.

To sum up, iterative reconstruction has several attractive features, although it is not as computationally efficient as analytical reconstruction. It allows the integration of various physical models, which can reduce noise and artifacts, depending

on modelling accuracy. It is also more forgiving of data truncation at a given view and view truncation as in limited view problems, leading to its biggest advantage, i.e. its great potential for radiation reduction. Despite all the advantages, there are some obstacles hindering the promotion of the iterative reconstruction technique, one of which is that it takes many years for the radiologist to adapt to the new noise patterns and artifacts. However, Beister points out that "A movement away from analytical methods in favor of iterative methods can nevertheless be anticipated for CT in general within the next few years." [10]

2.6 Conclusion

This chapter has introduced the fundamentals of computed tomography. First the development of CT was reviewed, starting from the invention of CT and followed by the illustration of different generations of scanners: parallel-beam, fan-beam and cone-beam. Then the physical theory of CT was described as well as the cause and appearance of various artifacts. In the following section, we focussed on low-radiation CT, revealing its benefits and guiding principles, the quantification of radiation and effective dose-management strategies. Finally, both the analytical and iterative reconstruction algorithms were depicted in detail, together with their advantages and disadvantages.

Chapter 3

Fundamentals of Compressed Sensing

3.1 Introduction

Sampling is the process of converting a signal, for example, a function of continuous time or space, into a numeric sequence, a function of discrete time or space. The intuitive expectation is that when the continuous function is reduced to discrete samples, one can still recover the original signal through interpolation. Kotelnikov, Nyquist, Shannon, and Whittaker tackled this problem by proposing the sampling theorem [86, 113, 135, 153]. Their pioneering work is a fundamental bridge between continuous signals and discrete signals, and is commonly known as the Nyquist-Shannon sampling theorem. It states that signals can be exactly recovered from a set of uniformly spaced samples taken at the Nyquist rate of twice the bandwidth of the signal of interest. This discovery has moved signal processing from the analogue to the digital domain, and digitization has enabled more robust, flexible and cheaper sensing and processing systems than their analogue counterparts [37]. However, with ever-increasing fidelity and resolution, the resulting Nyquist rate is so high that we have far too many samples, or in some cases, it is simply beyond the capability of physical devices. In X-ray computed tomography (CT), for example, the number of measurements required by the sampling theorem for a successful recovery is usually large and will eventually cause high dose of radiation. Hence, despite extraordinary advances in computa-

tional power, the acquisition and processing of signals such as images and videos continues to call for an even lower sampling rate.

On the other hand, compression techniques have emerged to deal with high-dimensional data, which aim at finding the most concise representation of a signal while allowing some acceptable distortion. They rely on the empirical observation that many types of signals can be well-approximated by a sparse expansion in terms of a suitable basis, that is, by only a small number of non-zero coefficients. Roughly speaking, one compresses the signal by simply keeping the largest coefficients when the signal is described by a basis or frame, which is known as transform coding. Many lossy compression techniques such as JPEG, JPEG2000, MPEG, and MP3 all comply with this principle. Leveraging the concept of transform coding, we are interested in whether there is a way to directly sense the data in a compressed form, rather than first sampling at a high rate and then compressing the sampled data.

The emergence of the novel theory of compressed sensing (CS) - also known as compressed sensing or compressive sampling - provides a satisfactory answer to the above question. The key feature of CS is to achieve perfect signal recovery from much fewer measurements than usually required by the Shannon-Nyquist criterion. CS exploits the sparseness or compressibility of signals in a predefined basis or frame, and its required number of measurements is proportional to the number of non-zero elements in the sparse representation of signals [24]. The technique of CS [63, 96] was proposed in 2006 with the publication of papers by Candès, Romberg, Tao [23] and by Donoho [40]. It has been a very active area in both theory and applications since then. The former group of authors introduced the restricted isometry property (RIP), which was initially called the uniform uncertainty principle and recognized as a key property of compressed sensing matrices. It has been proved that Gaussian, Bernoulli, and partial random Fourier matrices [21, 124, 131] possess this important property and randomness is exploited to show optimal or at least near-optimal conditions on the number of measurements in terms of the sparsity. On the other hand, Donoho approaches this problem via polytope geometry, more precisely, via the notion of k-neighbourliness. He observes sharp phase transition curves shown for Gaussian random matrices separating regions where recovery fails or succeeds with high probability [41–43]. The foundation of CS theory lies in probability theory, finite dimensional Banach space geometry [72, 90] and randomly-projected polytopes [4]. Interested readers

can refer to [51] for a historical overview.

There has been a notable impact of CS on several applications in imaging [46, 129], A/D conversion [147], radar [65] and wireless communication [119, 144], to name a few. One important area is medical imaging, where it has enabled speedy magnetic resonance (MR) scanning while preserving diagnostic quality [148]. Moreover, the broad applicability of this framework has inspired research that extends the CS framework to CT scanning for radiation reduction, which is in accordance with the theme of my thesis. The aim of this chapter is to provide an up-to-date review of some of the important results in CS from both theoretical and numerical aspects. We start with the introduction of the general mathematical model and the optimization algorithms in Section 3.2, followed by the two fundamental conditions, sparsity in Section 3.3 and solution uniqueness in Section 3.4. Finally, a brief review of the application of CS in CT reconstruction is given in Section 3.5.

3.2 Mathematical Model

We start with the preliminaries and notation. In what follows, $\mathbb{C}$, $\mathbb{R}$ denote the set of complex numbers and real numbers, respectively. I is the identity matrix, and $||\cdot||_p$ is the p-norm operator. For a vector object $\mathbf{x} \in \mathbb{C}^N$,

$$\begin{aligned} ||\mathbf{x}||_p &:= (\sum_{j=1}^{N} |x_j|^p)^{1/p}, 1 \leq p < \infty \\ ||\mathbf{x}||_0 &:= |\text{supp} \quad \mathbf{x}| \\ ||\mathbf{x}||_\infty &:= \max_{j=1,\cdots,N} |x_j| \end{aligned}$$

where supp $\mathbf{x}$ denotes the support of $\mathbf{x}$ and $|\cdot|$ stands for the cardinality. $\mathbf{x}$ is k-sparse if $||\mathbf{x}||_0 \leq k$. Let $\sum_k$ denote the set of k-sparse vectors, inf the infimum or greatest lower bound of a subset of real numbers. The best k-term approximation error of a vector $\mathbf{x} \in \mathbb{C}^N$ in ℓ_p is defined as

$$\sigma_k(\mathbf{x})_p = \inf_{\mathbf{z} \in \sum_k} ||\mathbf{x} - \mathbf{z}||_p.$$

For a matrix $A = (a_{ij}) \in \mathbb{C}^{m \times N}$, the ℓ_2 norm and Frobenius norm is defined as:

$$\begin{aligned} ||A||_2 &= \sigma_{\max}(A) \\ ||A||_F &= \sqrt{\sum_i \sum_j |a_{ij}|^2} \end{aligned}$$

where $\sigma_{\max}(A)$ denotes the largest singular value of A. In addition, it is known that

$$||A||_2 \leq ||A||_F.$$

For a subset $\mathbf{T} \subset \{1, \cdots, N\}$ and its complement $\mathbf{T}^c = \{1, \cdots, N\} \setminus \mathbf{T}$, $\mathbf{x_T}$ stands for the vector that only takes the entries of $\mathbf{x}$ in $\mathbf{T}$ and sets the entries 0 outside $\mathbf{T}$. Similarly, $A_\mathbf{T}$ denotes the column sub-matrix of A consisting of columns indexed by $\mathbf{T}$.

This completes the notation.

Suppose the measurement vector $\mathbf{b} \in \mathbb{C}^m$ of a signal $\mathbf{x} \in \mathbb{C}^N$ is obtained via:

$$A\mathbf{x} = \mathbf{b} \tag{3.1}$$

where $A \in \mathbb{C}^{m \times N}$ is the measurement/sensing matrix. The main interest is in the heavily under-sampled case $m \ll N$, where it would be impossible to recover $\mathbf{x}$ from $\mathbf{b}$ without any further information. However, the additional assumption that $\mathbf{x}$ is k-sparse can change the situation dramatically. It naturally leads to the ℓ_0-minimization problem:

$$\min \quad ||\mathbf{x}||_0 \quad \text{subject to} \quad A\mathbf{x} = \mathbf{b} \tag{3.2}$$

where $||\mathbf{x}||_0$ denotes the ℓ_0 norm that counts the number of nonzero entries of $\mathbf{x}$. Unfortunately, the ℓ_0-minimization problem is non-deterministic polynomial-time (NP) hard [98,105] and therefore computationally intractable. Two practical alternatives to (3.2) have been proposed: basis pursuit [30] and greedy algorithms [106].

Basis pursuit algorithms

This group of algorithms relaxes $||\cdot||_0$ to $||\cdot||_1$, so the problem is converted

to convex optimization as in (3.3), for which there exist efficient and accurate numerical solvers.

$$\min \quad ||\mathbf{x}||_1 \quad \text{subject to} \quad A\mathbf{x} = \mathbf{b} \tag{3.3}$$

where $||\mathbf{x}||_1$ denotes the ℓ_1 norm which is the sum of the absolute entries of $\mathbf{x}$. It may not be immediately clear that the solution to (3.3) will be at all similar to the solution to (3.2). Therefore Figure 3.1 is given to explain that the use of ℓ_1 promotes sparsity.

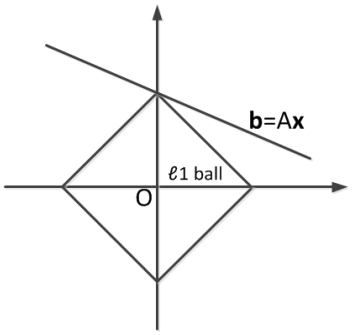


Figure 3.1: The ℓ_1-minimizer within the affine space of solutions of the linear system $A\mathbf{x} = \mathbf{b}$ coincides with a sparsest solution

In the presence of noise, the linear program in (3.3) is relaxed to the convex program with a conic constraint:

$$\min \quad ||\mathbf{x}||_1 \quad \text{subject to} \quad ||A\mathbf{x} - \mathbf{b}||_2 \leq \epsilon$$

or the unconstrained version:

$$\min \quad \frac{\tau}{2}||A\mathbf{x} - \mathbf{b}||_2 + ||\mathbf{x}||_1$$

where τ and ϵ are both parameters, and with appropriate choice, these two problems are equivalent. There are a great variety of algorithms designed to explicitly solve these problems in the context of CS [9, 50, 52, 93, 115, 152, 162].

Greedy algorithms

Greedy algorithms rely on iterative approximation of the signal coefficients and support, either by iteratively identifying the support of the signal until a convergence criterion is met, or alternatively by obtaining an improved estimate of the

sparse signal at each iteration that attempts to account for the mismatch to the measured data [37]. Two of the oldest and simplest greedy approaches are orthogonal matching pursuit (OMP) [36, 44, 106] and iterative thresholding [15]. OMP seeks a sparse minimizer one entry at a time. It starts with finding the column of A most correlated to the measurements, then repeats this step by correlating the columns with the signal residual, which is obtained by subtracting the contribution of a partial estimate of the signal from the original measurement vector. Iterative thresholding algorithms are more straightforward, and iterate a gradient descent step followed by thresholding until a convergence criterion is met. There have been many efforts to improve upon these basic results [35, 39, 107, 108].

3.3 Sparse and Compressible Signals

The great potential of CS in many applications is on the grounds that a wide range of real-world signals and images have a sparse or compressible expansion in terms of a suitable basis or frame, for example a wavelet expansion. We have introduced the definition of sparse signals in Section 3.2. However, the signal of practical interest may not be strictly sparse in space or in a transform domain, but take the form of rapid decay, typically like a power law. Such type of signals is known as compressible.

We now consider objects whose coefficients on some basis decay like a power-law [21]. A finite signal $\mathbf{x} \in \mathbb{C}^N$ can be described by an orthonormal basis φ as $\varphi(\mathbf{x})$, which can be arranged in decreasing order of magnitude $|\varphi|_{(1)} \geq |\varphi|_{(2)} \geq \cdots \geq |\varphi|_{(N)}$. $\varphi(\mathbf{x})$ is said to belong to the weak-ℓ_p ball of radius B for some $0 < p < \infty$ if for each $i \in [1 \quad N]$,

$$|\varphi|_{(i)} \leq B \cdot i^{-1/p}. \tag{3.4}$$

p controls the speed of the decay: the smaller p, the faster the decay. It is well-known that the decay rate of the coefficients of $\mathbf{x}$ is linked to the 'compressibility' of $\mathbf{x}$. Suppose $\varphi_k(\mathbf{x})(1 \leq k \leq N)$ is obtained by keeping the k largest entries of the vector $\varphi(\mathbf{x})$ and setting the others to zero. Then it immediately follows from (3.4) that the approximation error obeys

$$||\varphi(\mathbf{x}) - \varphi_k(\mathbf{x})||_2 \leq C_p \cdot B \cdot k^{-t}, t = 1/p - 1/2,$$

for some constant C_p only dependent on p. Hence, it follows from Parseval that the approximate signal $\mathbf{x}_k$ reconstructed from $\varphi(\mathbf{x})$ obeys the same estimate,

$$||\mathbf{x} - \mathbf{x}_k||_2 \leq C_p \cdot B \cdot k^{-t}. \tag{3.5}$$

CS states if one makes some random measurements of a signal, and then reconstructs an approximate signal from this limited set of measurements in a manner which requires no prior knowledge of or assumptions about the signal except that it obeys power-law decay in some domain, one still obtains a reconstruction error which is equally as good as that one would obtain by knowing everything about the signal and selecting the k largest entries of the coefficient vector in that domain.

3.4 Uniqueness of k-sparse Solutions

To guarantee successful recovery with CS, the condition must be established that there is only one k-sparse solution to $A\mathbf{x} = \mathbf{b}$. In this section, two fundamental properties of the measurement matrix are introduced, and the definitions, lemmas and theorems below are quoted from [51].

3.4.1 Null Space Property

The first property is called the null space property (NSP). The null space of the measurement matrix is denoted as $\mathcal{N}(A) = \{\mathbf{x} : A\mathbf{x} = 0, \mathbf{x} \in \mathbb{C}^N\}$ [19].

Definition 3.1. *A matrix $A \in \mathbb{C}^{m \times N}$ satisfies the NSP of order k with constant $\alpha \in (0, 1)$ if*

$$||\mathbf{x}_\mathbf{T}||_1 \leq \alpha ||\mathbf{x}_{\mathbf{T}^c}||_1$$

for all sets $\mathbf{T} \subset \{1, \cdots, N\}$, $|\mathbf{T}| \leq k$ and for all $\mathbf{x} \in \mathcal{N}(A)$.

With this notion, we have the following theorem.

Theorem 3.2. *[51] Let $A \in \mathbb{C}^{m \times N}$ be a matrix that satisfies the NSP of order k with constant $\alpha \in (0, 1)$, $\mathbf{x} \in \mathbb{C}^N$ and $\mathbf{b} = A\mathbf{x}$, $\mathbf{x}^*$ be a solution of the ℓ_1-*

minimization problem (3.3). *Then*

$$||\mathbf{x} - \mathbf{x}^*||_1 \leq \frac{2(1+\alpha)}{1-\alpha}\sigma_k(x)_1.$$

In particular, if $\mathbf{x}$ *is* k*-sparse,* $\mathbf{x}^* = \mathbf{x}$.

Proof. Since x^* is a solution of the ℓ_1-minimization problem (3.3), we have:

$$||\mathbf{x}^*||_1 \leq ||\mathbf{x}||_1.$$

Let $\mathbf{T}$ be the set of the k-largest entries of $\mathbf{x}$ in absolute value, then

$$||\mathbf{x}^*_{\mathbf{T}}||_1 + ||\mathbf{x}^*_{\mathbf{T}^c}||_1 \leq ||\mathbf{x}_{\mathbf{T}}||_1 + ||\mathbf{x}_{\mathbf{T}^c}||_1. \tag{3.6}$$

Let $\boldsymbol{\eta} = \mathbf{x}^* - \mathbf{x}$, then $\boldsymbol{\eta} \in \mathcal{N}(A)$. Because of triangle inequality, which is

$$||\mathbf{x}_{\mathbf{T}}||_1 - ||\boldsymbol{\eta}_{\mathbf{T}}||_1 \leq ||\mathbf{x}^*_{\mathbf{T}}||_1, ||\boldsymbol{\eta}_{\mathbf{T}^c}||_1 - ||\mathbf{x}_{\mathbf{T}^c}||_1 \leq ||\mathbf{x}^*_{\mathbf{T}^c}||_1$$

(3.6) becomes:

$$\begin{gathered} ||\mathbf{x}_{\mathbf{T}}||_1 - ||\boldsymbol{\eta}_{\mathbf{T}}||_1 + ||\boldsymbol{\eta}_{\mathbf{T}^c}||_1 - ||\mathbf{x}_{\mathbf{T}^c}||_1 \leq ||\mathbf{x}_{\mathbf{T}}||_1 + ||\mathbf{x}_{\mathbf{T}^c}||_1 \\ ||\boldsymbol{\eta}_{\mathbf{T}^c}||_1 \leq ||\boldsymbol{\eta}_{\mathbf{T}}||_1 + 2||\mathbf{x}_{\mathbf{T}^c}||_1 \leq \alpha||\boldsymbol{\eta}_{\mathbf{T}^c}||_1 + 2\sigma_k(\mathbf{x})_1 \\ ||\boldsymbol{\eta}_{\mathbf{T}^c}||_1 \leq \tfrac{2}{1-\alpha}\sigma_k(\mathbf{x})_1 \end{gathered} \tag{3.7}$$

Hence,

$$||\mathbf{x} - \mathbf{x}^*||_1 = ||\boldsymbol{\eta}_{\boldsymbol{T}}||_1 + ||\boldsymbol{\eta}_{\mathbf{T}^c}||_1 \leq (\alpha+1)||\boldsymbol{\eta}_{\mathbf{T}^c}||_1 \leq \frac{2(1+\alpha)}{1-\alpha}\sigma_k(x)_1$$

This completes the proof. □

3.4.2 Restricted Isometry Property

The second property is called the restricted isometry property (RIP).

Definition 3.3. *An* $m \times N$ *matrix* $A \in \mathbb{C}^{m\times N}$ *satisfies the RIP of order* k *if there is some constant* $\delta_k \in (0,1)$ *such that*

$$(1-\delta_k)||\mathbf{x}||_2^2 \leq ||A\mathbf{x}||_2^2 \leq (1+\delta_k)||\mathbf{x}||_2^2 \tag{3.8}$$

holds for all k-sparse vectors $\mathbf{x} \in \mathbb{C}^N$. δ_k *here is defined as the restricted isometry constant.*

When $\delta_k \in (0,1)$, A satisfies the RIP of order k with constant δ_k. The restricted isometry constant δ_k can also be equivalently defined as:

$$\delta_k = \max_{\mathbf{T} \subset \{1,\cdots,N\}, |\mathbf{T}| \leq k} ||A_{\mathbf{T}}^* A_{\mathbf{T}} - I||_2$$

The relationship between RIP and NSP is revealed in the lemma below.

Lemma 3.4. *[51] Assume that* $A \in \mathbb{C}^{m \times N}$ *satisfies RIP of order* $K = k + h$ *with constant* $\delta_K \in (0,1)$. *Then A has the NSP of order k with constant* $\alpha = \sqrt{\frac{k(1+\delta_K)}{h(1+\delta_K)}}$.

Proof. From the assumption, we have:

$$(1-\delta_K)||\mathbf{x}||_2^2 \leq ||A\mathbf{x}||_2^2 \leq (1+\delta_K)||\mathbf{x}||_2^2, \quad \forall \mathbf{x} \in \sum_K \tag{3.9}$$

Let $\boldsymbol{\eta} \in \mathcal{N}(A)$, whose entry indexes are divided to disjoint sets $\mathbf{T}_0, \mathbf{T}_1, \cdots, \mathbf{T}_q$. $\mathbf{T}_0 = \mathbf{T} \subset \{1, \cdots, N\}, |T| \leq k$, and the size of other sets, i.e. $\mathbf{T}_1, \cdots, \mathbf{T}_q$, is at most h. In addition, the entries are arranged in a non-increasing order on $\mathbf{T}_1, \cdots, \mathbf{T}_q$ as below:

$$|\eta_j| \leq |\eta_i| \quad \text{for all} \quad j \in \mathbf{T}_p, i \in \mathbf{T}_{p'}, 0 \leq p' \leq p \leq q$$

Note that $A\boldsymbol{\eta} = 0$ implies $A\boldsymbol{\eta}_{\mathbf{T}_0 \cup T_1} + \sum_{j=2}^{q} A\boldsymbol{\eta}_{\mathbf{T}_j} = 0$.

Based on the Cauchy-Schwarz inequality, RIP and the triangle inequality, the following sequence of inequalities is deduced:

$$\begin{aligned}
||\boldsymbol{\eta}_{\mathbf{T}}||_1 &\leq \sqrt{k}||\boldsymbol{\eta}_{\mathbf{T}}||_2 \leq \sqrt{k}||\boldsymbol{\eta}_{\mathbf{T}_0 \cup \mathbf{T}_1}||_2 \\
\text{from (3.9)} &\leq \sqrt{\frac{k}{1-\delta_K}}||A\boldsymbol{\eta}_{\mathbf{T}_0 \cup \mathbf{T}_1}||_2 \\
&= \sqrt{\frac{k}{1-\delta_K}}||A\boldsymbol{\eta}_{\mathbf{T}_2 \cup \mathbf{T}_3 \cup \cdots \cup \mathbf{T}_q}||_2 \\
&\leq \sqrt{\frac{k}{1-\delta_K}} \sum_{j=2}^{q} ||A\boldsymbol{\eta}_{\mathbf{T}_j}||_2
\end{aligned}$$

$$\begin{aligned}
\text{from (3.9)} &\leq \sqrt{\frac{1+\delta_K}{1-\delta_K}}\sqrt{k}\sum_{j=2}^{q}||\boldsymbol{\eta}_{\mathbf{T}_j}||_2 \\
\text{from the non-increasing arrangement} &\leq \sqrt{\frac{1+\delta_K}{1-\delta_K}}\frac{k}{h}\sum_{j=1}^{q-1}||\boldsymbol{\eta}_{\mathbf{T}_j}||_1 \\
&\leq \sqrt{\frac{1+\delta_K}{1-\delta_K}}\frac{k}{h}||\boldsymbol{\eta}_{\mathbf{T}^c}||_1
\end{aligned}$$

This concludes the proof. □

Taking $h = 2k$ and $\sqrt{\frac{1+\delta_K}{1-\delta_K}\frac{k}{h}} < 1$ leads to the conclusion that A satisfies the NSP of order k when $\delta_{3k} < \frac{1}{3}$. By Theorem 3.2, the stable recovery of the k-sparse vector by ℓ_1-minimization is guaranteed. It is shown in the following theorem that RIP also puts a bound on the reconstruction error in ℓ_2.

Theorem 3.5. *[51] that $A \in \mathbf{C}^{m\times N}$ satisfies RIP of order $3k$ with constant $\delta_{3k} \leq \frac{1}{3}$. For $\mathbf{x} \in \mathbf{C}^N$, let $\mathbf{b} = A\mathbf{x}$ and x^* be the solution of the ℓ_1-minimization problem* (3.3). *Then*

$$||\mathbf{x} - \mathbf{x}^*||_2 \leq C\frac{\sigma_k(x)_1}{\sqrt{k}}$$

with $C = \frac{2}{1-\alpha}(\frac{\alpha+1}{\sqrt{2}} + \alpha), \alpha = \sqrt{\frac{1+\delta_{3k}}{2(1-\delta_{3k})}}$

Proof. Based on the proof of Lemma 3.4, the similar setting is adopted with $\boldsymbol{\eta} = \mathbf{x}^* - \mathbf{x} \in \mathcal{N}(A)$, $h = 2k$ and we have:

$$||\boldsymbol{\eta}_{\mathbf{T}}||_2 \leq \sqrt{\frac{1+\delta_{3k}}{2k(1-\delta_{3k})}}||\boldsymbol{\eta}_{\mathbf{T}^c}||_1.$$

Since $\delta_{3k} \leq \frac{1}{3}$ makes $\alpha < 1$, we have:

$$\begin{aligned}
||\boldsymbol{\eta}_{\mathbf{T}^c}||_2 &\leq \sqrt{\frac{1}{2k}}||\boldsymbol{\eta}||_1 = \sqrt{\frac{1}{2k}}(||\boldsymbol{\eta}_{\mathbf{T}}||_1 + ||\boldsymbol{\eta}_{\mathbf{T}^c}||_1 \\
\text{from Definition 3.1} &\leq \sqrt{\frac{1}{2k}}(\alpha||\boldsymbol{\eta}_{\mathbf{T}^c}||_1 + ||\boldsymbol{\eta}_{\mathbf{T}^c}||_1 = \frac{\alpha+1}{\sqrt{2k}}||\boldsymbol{\eta}_{\mathbf{T}^c}||_1
\end{aligned}$$

Finally,

$$\begin{aligned}
||\mathbf{x} - \mathbf{x}^*||_2 &\leq ||\boldsymbol{\eta}_{\mathbf{T}}||_2 + ||\boldsymbol{\eta}_{\mathbf{T}^c}||_2 \\
&\leq (\frac{\alpha+1}{\sqrt{2}} + \alpha)\frac{||\boldsymbol{\eta}_{\mathbf{T}^c}||_1}{\sqrt{k}}
\end{aligned}$$

$$\text{from (3.7)} \quad \leq \quad \frac{2}{1-\alpha}\left(\frac{\alpha+1}{\sqrt{2}}+\alpha\right)\frac{\sigma_k(\mathbf{x})_1}{\sqrt{k}}$$

This completes the proof. □

Although the RIP itself is no easier to verify than the null space property, its advantage is the fact that it can be shown to hold with high probability for large classes of matrices generated by certain random procedures, such as Gaussian, Bernoulli and partial Fourier random matrices.

3.4.3 RIP for Special Types of Random Matrices

Optimal estimates for the RIP constants of measurement matrices in terms of the sample number m can be obtained for Gaussian, Bernoulli or more general sub-Gaussian random matrices. The entries of Gaussian matrices are i.i.d. Gaussian random variables with expectation 0 and variance $1/m$, and the entries of Bernoulli matrices take the value $1/\sqrt{m}$ or $-1/\sqrt{m}$ with equal probability. Both satisfy the concentration inequality:

$$P(|||A\mathbf{x}||_2^2 - ||\mathbf{x}||_2^2| \geq \delta||\mathbf{x}||_2^2) \leq 2e^{-c_0\delta^2 m}, \delta \in (0,1), \tag{3.10}$$

with $c_0 > 0$ is some constant. Correspondingly, their RIP is based on this inequality.

Theorem 3.6. *[51] Let $A \in \mathbb{R}^{m\times N}$ be a random matrix satisfying the concentration property* (3.10)*, there exists a constant C depending only on c_0 such that the restricted isometry constant of A satisfies $\delta_k \leq \delta$ with probability at least $1-\varepsilon$ provided*

$$m \geq C\delta^{-2}(k\log(N/M) + \log(\varepsilon^{-1})), \varepsilon \in (0,1).$$

Random partial Fourier matrix is another important class regarding the application of CS. Compared with Gaussian and Bernoulli matrix, it tends to be more justifiable in practice, as well as more efficient, due to the availability of speedy matrix-vector multiplication via fast Fourier transform (FFT). It was also the study object of the very first papers on CS [21, 23]. A random partial Fourier matrix $A \in \mathbb{C}^{m\times N}$ is obtained from the discrete Fourier matrix $F \in \mathbb{C}^{N\times N}$ with

entries

$$F_{p,q} = \frac{1}{\sqrt{N}} e^{-j2\pi pq/N}, \quad p, q = 0, 1, \cdots, N-1$$

by selecting m rows uniformly at random among the full N rows.

To start with, we present two examples in Figure 3.2 and Figure 3.3 for some intuitive understanding. The Fourier coefficients of the signal samples gained via different ways are plotted. It is noticed that when the sampling ratio is lower than that required by the Nyquist-Shannon theorem, random sampling in time/space domain is able to preserve the frequency information better than uniform sampling.

The following theorem concerning the RIP of random partial Fourier matrices is proven in [125], and is improved slightly in [21, 124, 131].

Theorem 3.7. *Let $A \in \mathbb{C}^{m \times N}$ be the random partial Fourier matrix, the restricted isometry constant of the rescaled matrix $\sqrt{\frac{N}{m}} A$ satisfies $\delta_k \leq \delta$ with probability exceeding $1 - N^{-\alpha \log^3(N)}$ provided*

$$m \geq C \delta^{-2} k \log^4(N)$$

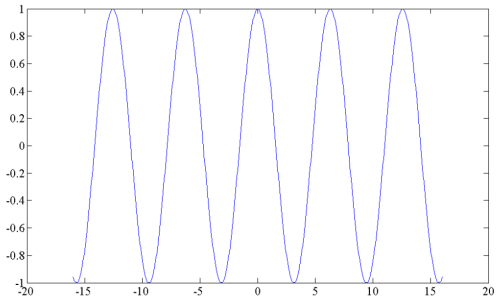
where $C, \alpha > 1$ are universal constants.

3.5 CS-based CT reconstruction

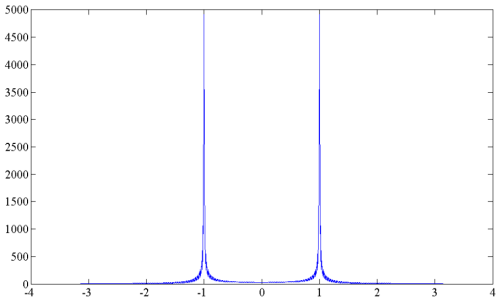
CS theory has been exploited in many applications since it was born, among which medical imaging is a key area. The mathematical framework of CS is elegant and compatible with the CT iterative reconstruction technique, formulated as:

$$\min \quad ||\varphi \mathbf{x}||_1 \quad s.t. \quad A\mathbf{x} = \mathbf{b} \tag{3.11}$$

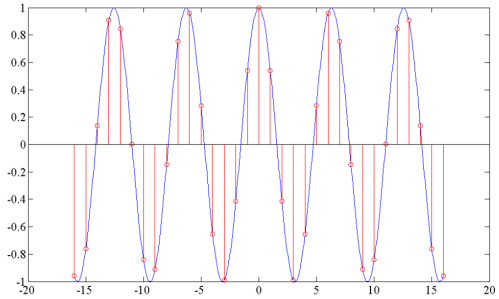
where $\mathbf{x}$ is the object to be reconstructed, A is the CT system matrix, or the projection operator that converts the object to projections, $\mathbf{b}$ is the measurement, and φ is the sparsifying basis. One may view it as simply another plain iterative algorithm, but the publication of CS has caused a dramatic revitalization in the image reconstruction community [143]. What makes the CS theory so important is that it addresses the sampling condition for accurate reconstruction of an image object, to be specific, the sparse property and the RIP.



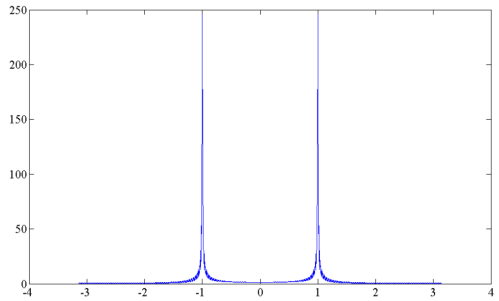
(a) *Original signal*



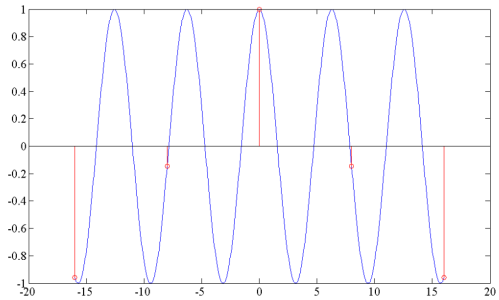
(b) *Fourier transform of the original signal*



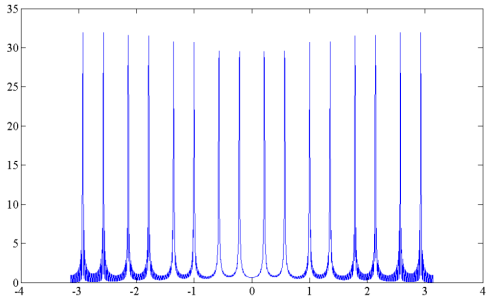
(c) *Fast regular sampling that satisfies Nyquist-Shannon theorem*



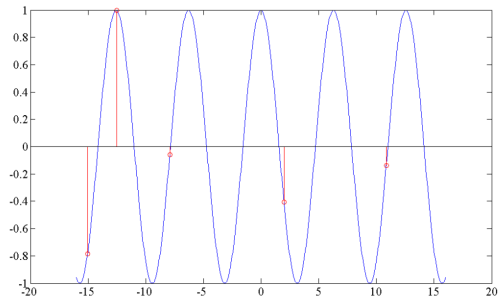
(d) *Fourier transform of the samples gained via fast regular sampling*



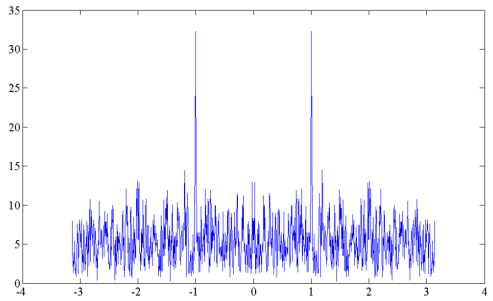
(e) *Slow regular sampling*



(f) *Fourier transform of the samples gained via slow regular sampling*

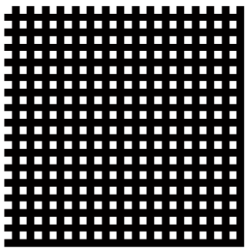


(g) *Random sampling*

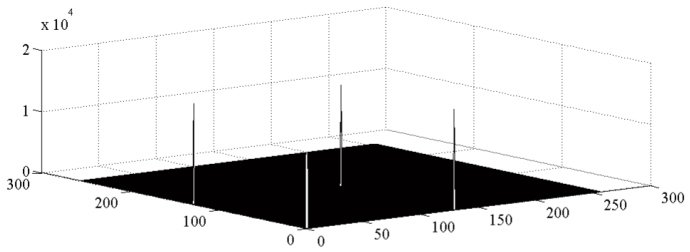


(h) *Fourier transform of the random samples*

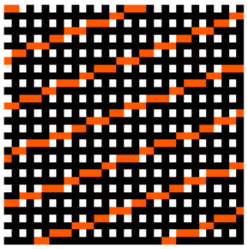
Figure 3.2: Comparison of Fourier coefficients of 1-D signal samples gained via different sampling patterns



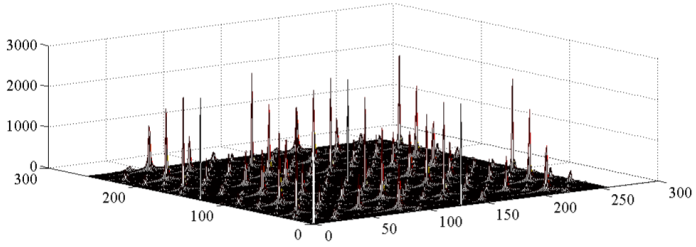
(a) *Dense uniform samples*



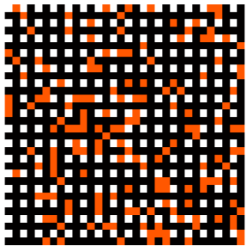
(b) *Fourier transform of dense uniform samples*



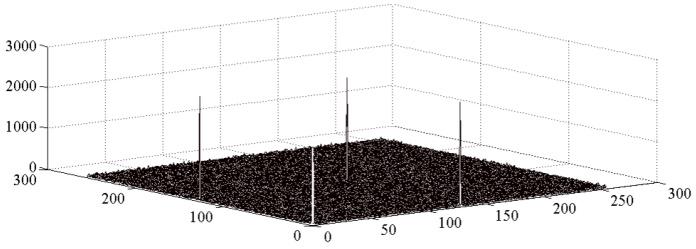
(c) *Sparse uniform samples*



(d) *Fourier transform of sparse uniform samples*



(e) *Sparse random samples*



(f) *Fourier transform of sparse random samples*

Figure 3.3: Comparison of Fourier coefficients of 2-D image samples gained via different sampling patterns

Here we review briefly the application of CS in CT reconstruction. Sidky and Pan et al. started to apply CS to CT reconstruction in 2006 [139], using total variation(TV) minimization and incomplete projection data. This model has been particularly useful and has been adapted by many researchers since then [13, 33, 128, 128, 140, 141, 143, 163]. The main innovations concern three aspects: (1) computationally efficient parallel programming with appropriate hardware and (2) mathematical formulation of an efficient optimization algorithm for fast convergence and (3) incorporation of other constraints for higher reconstruction accuracy. The first issue is due to the iterative nature in solving the model; in general, since it requires multiple iterations of forward and backward projections of large datasets, it cannot be completed in a clinically feasible time frame (e.g.,

<1 min). The problem is resolved successfully with the use of graphics processing units (GPUs) [69, 70], which can reduce the computational time from several hours to a few minutes. For the second issue, a wide range of algorithms has been exploited, including Newton's method, augmented Lagrangian method [91], split Bregman method [56], Barzilai-Borwein formulation [116], etc. The third issue involves other constraints, e.g. minimization of the ℓ_1 norm of the signal in wavelet domain [68], or formulating the CS method in the framework of a different model, e.g. statistical reconstruction [143].

A natural and important extension of the TV-based method is referred to as the prior image constrained compressed sensing (PICCS) proposed by Chen et al. [29]. They exploit the application of CS in dynamic CT imaging. A prior image $\mathbf{x}_p$ reconstructed from the union of interleaved dynamical data sets is utilized to sparsify dynamic CT image sequences, which is favourable for CS image reconstruction for individual time frames. Let φ_1 and φ_2 denote the sparsifying transform, α the parameter. The reconstruction problem is formulated as:

$$\min \quad \alpha||\varphi_1(\mathbf{x}-\mathbf{x}_p)||_1 + (1-\alpha)||\varphi_2\mathbf{x}||_1 \quad s.t. \quad A\mathbf{x} = \mathbf{b} \tag{3.12}$$

The shortcoming of PICCS is that it assumes that the prior images and the current images are taken at the same global geometrical coordinates, and the pixels with the same attenuation values correspond to the same scanning data [109]. This assumption, however, does not necessarily translate into practical situation due to inevitable error and noise. Various alternative methods of PICCS have been proposed in the use of different prior knowledge, such as image intensity histograms [123] and image features [154].

3.6 Conclusion

In this chapter, we have laid the foundation of compressed sensing. To begin with, we made the problem statement and gave the associated two main solvers: basis pursuit and greedy algorithms. Then the two key elements of CS theorem were described in detail: sparse/compressible signals and the RIP of the sensing matrix. These two work together to guarantee the uniqueness of the solution. RIP for random Fourier matrix was specifically mentioned as it is related closely to the CT system, as will be shown in the next chapter. Finally, we presented

the application of CS in CT, highlighting TV regularization and PICSS.

Chapter 4

Analysis of Sampling Schemes in CT Scans

4.1 Introduction

As introduced in Chapter 2, the significant clinical benefits of computed tomography (CT) in health care were immediately recognized following its introduction into clinical practice in 1972, when Sir Godfrey Hounseld at EMI patented the first CT scanner. Since it was invented, this imaging technology has seen extensive growth in the number of CT scanners and the frequency of CT examinations. The credit for its popularity depends on its capability of providing a wide coverage of the area of interest in a short time, and its high spatial resolution and ease of operation. Hence, as a popular and useful medical tool, CT plays an important role in diagnosis and thus benefits patients greatly. However, CT scanning contributes over 44% to the total collective effective dose equivalent from global medical exposures. It is known that X-ray radiation leads to ionization of body cells and increased radiation dose raises the risk of cancer, which makes the reduction of X-ray radiation dose delivered to patients no trivial task. Of the various radiation management strategies, to under-sample the projections is quite straightforward as the under-sampling factor implies the same reduction of potential radiation dose. We describe different sampling schemes in this chapter, and investigate their impacts on the Fourier slice theorem (FST)-based projection model.

The rest of the chapter is organized as follows. Section 4.2 introduces different sampling schemes in fan-beam CT. Then the projection model based on FST is built in Section 4.3 and verified by simulations in Section 4.5. The analysis of the impacts of different schemes on the projection model is made in Section 4.4. Finally, conclusions are drawn in Section 4.6.

4.2 Sampling Schemes

The acquisition of ordinary fan-beam CT is illustrated in Figure 4.1, and the projection data is measured from $1000 - 2000$ X-ray source positions uniformly distributed over the angular range $[0 \quad 2\pi)$. For the mathematical analysis throughout Chapters 3 and 4, we refer to the scanning data with $2n$ views and $2n$ detectors as the full data for a $n \times n$ object, as proposed in [73] .

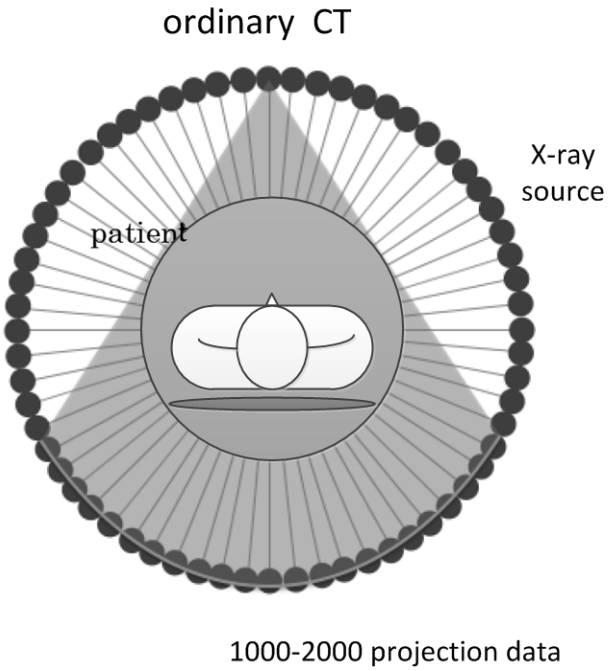


Figure 4.1: Schematic illustration of data acquisition in ordinary CT

In order to take fewer projections, one possible and popular approach is the sparse-view sampling scheme [3,13,62], which was applied earlier in CT perfusion of brains to see the temporal changes in the images. Besides radiation reduction, the use of fewer projection views can also shorten imaging time in CT scanning, improve work flow and minimize potential motion artifacts. The principle of sparse-view CT is shown in Figure 4.2, in which the number of projection views

is reduced to less than 100, as determined by the application and the required image quality. The challenging part of this scheme lies in its requirement for a fast on-off switching of the tube power, but even that development may not be far off due to the continuous efforts of manufacturers. Depending on the distribution of the X-ray source, there are two possible schemes: golden-angle and random-angle. As the name suggests, the entries of locations are distributed either uniformly or randomly over the angular range $[0 \quad 2\pi)$. In the golden-angle scheme [26], radial projections are successively incremented by the golden angle $\frac{\sqrt{5}-1}{2} \cdot 180^o \approx 111.25^o$, which can generate the sets of projections with relatively uniform angular spacings, regardless of the number of radial projections in the set.

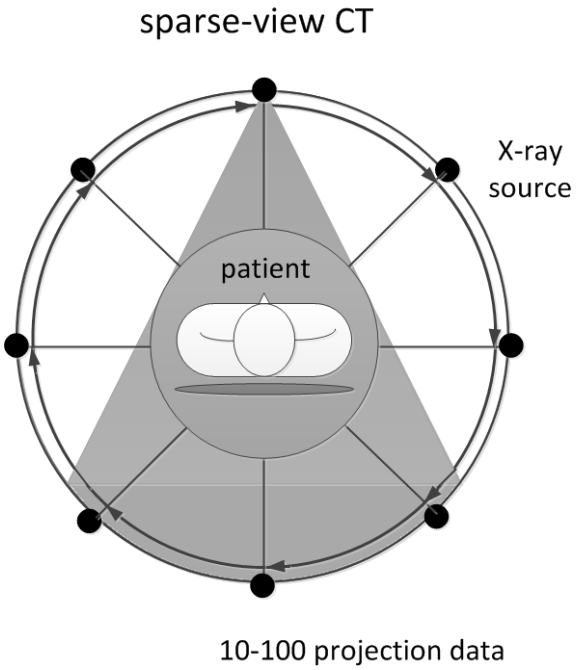


Figure 4.2: Schematic illustration of data acquisition in sparse-view CT

The other realization is the sparse-detector sampling scheme proposed by Cho and Lee et al. [2,32], which uses an oscillating multi-slit collimator between the X-ray source and the patient, as shown in Figure 4.3. Here the collimator during a scan can be reciprocated. This efficiently reduces the radiation dose to the patient by substantially blocking the radiation. Similar to sparse-view CT, the distribution of detectors can also be uniform or random, known as regular-detector and random-detector sampling schemes, respectively. To increase the incoherence of the scanning scheme, the collimator is reciprocated to achieve different detector distributions at every view in the random-detector sampling scheme.

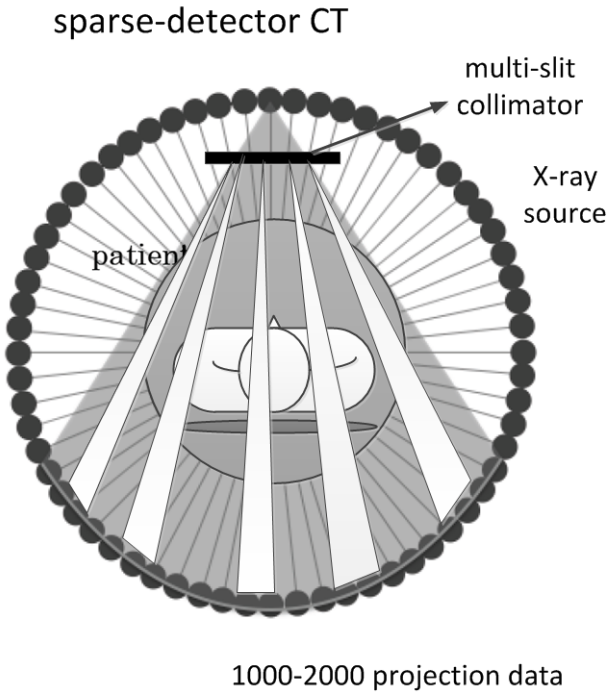


Figure 4.3: Schematic illustration of data acquisition in sparse-detector CT

Using the same setting as in Figure 2.8 and suppose the scanning comes with q views and p detectors, we can express the four sampling schemes mathematically. In sparse-view scheme, the source location $[x_s \quad y_s]$ is denoted by:

$$[x_s \quad y_s] = [R\cos\beta \quad R\sin\beta]\delta(\beta - \beta_0)$$

with $\beta_0 = \mod \{[1, 2, \cdots, q] \cdot 111.25^o, 360^o\}$ for the golden-angle scheme,

and β_0 are q uniformly distributed random variables from $[0 \quad 360^o)$

for the random-angle scheme.

In sparse-detector scheme, the detector distribution within a fan is denoted by $\delta(\gamma - \gamma_0)$, with

$$\gamma_0 = \frac{2\gamma_m}{p-1}[\frac{1-p}{2}, \frac{1-p}{2} + 1, \cdots, 0, \cdots, \frac{p-1}{2} - 1, \frac{p-1}{2}]$$

for the regular-detector scheme, and γ_0 are p uniformly distributed random variables from $[-\gamma_m \quad \gamma_m]$ for the random-detector scheme.

To sum up, we have two main types of under-sampling: sparse-view and sparse-detector. Each type has two realizations, leading to four specific schemes: golden-

angle and random-angle, and regular-detector and random-detector. In the next sections, we build the projection model and see how these four schemes affect the model.

4.3 Projection Model Based on Fourier Slice Theorem

4.3.1 Relationship between Different Fourier Transforms

The derivation details of Fourier slice theorem are revealed in Section 2.4.1. FST holds in the continuous-time domain, but the calculation in practice is carried out in the discrete-time domain. This gives rise to the need to study the relationship between different Fourier transforms [161].

The continuous-time Fourier transform (CTFT) is defined by the following pair of equations:

$$\begin{aligned} G_c(\Omega) &= \int_{-\infty}^{\infty} g_c(t) e^{-j\Omega t} dt \\ g_c(t) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} G_c(\Omega) e^{j\Omega t} d\Omega \end{aligned}$$

where $g_c(t)$ is the continuous-time signal and $G_c(\Omega)$ is its CTFT.

The discrete-time Fourier transform (DTFT) is defined by:

$$\begin{aligned} G_d(w) &= \sum_{l=-\infty}^{\infty} g_d[l] e^{-jwl} \\ g_d[l] &= \frac{1}{2\pi} \int_{-\pi}^{\pi} X(w) e^{jwl} dw \end{aligned}$$

where $g_d[l]$ is the discrete-time version of $g_c(t)$, i.e. $g_d[l] = g_c(lT)$ with T as the sampling period, and $G_d(w)$ is its DTFT.

As a bridge between $g_d[l]$ and $g_c(t)$, we consider a continuous-time signal that is

equivalent to $g_d[l]$ and is the sampled version of $g_c(t)$:

$$g(t) = g_c(t) \sum_{l=-\infty}^{\infty} \delta(t - lT)$$

The CTFT of $g(t)$ is:

$$\begin{aligned} G(\Omega) &= \int_{-\infty}^{\infty} g(t)e^{-j\Omega t}dt = \int_{-\infty}^{\infty} g_c(t) \sum_{l=-\infty}^{\infty} \delta(t - lT)e^{-j\Omega t}dt \\ &= \sum_{l=-\infty}^{\infty} \int_{-\infty}^{\infty} g_c(t)e^{-j\Omega t}\delta(t - lT)dt = \sum_{l=-\infty}^{\infty} g_c(lT)e^{-j\Omega lT} \\ &= \sum_{l=-\infty}^{\infty} g_d[l]e^{-jwl}|_{w=\Omega T} = G_d(w) \end{aligned} \tag{4.1}$$

On the other hand, the relationship between $G(\Omega)$ and $G_c(\Omega)$ can be revealed based on the product property of CTFT:

$$G(\Omega) = \frac{1}{T} \sum_{q=-\infty}^{\infty} G_c(\Omega + \frac{2\pi q}{T}) \tag{4.2}$$

With the above (4.1) and (4.2), we find the relationship between $G_c(\Omega)$ and $G_d(w)$:

$$G_d(w) = \frac{1}{T} \sum_{q=-\infty}^{\infty} G_c(\frac{w}{T} + \frac{2\pi q}{T}) \tag{4.3}$$

In other words, the DTFT is simply the scaled sum of shifted versions of the original CTFT.

The DTFT is computationally advantageous over CTFT, but is still not convenient since it is continuous in the frequency domain and thus requires an integration for inverse DTFT. That is why DTFT is sampled in the frequency domain to make the discrete Fourier transform (DFT), which is defined as:

$$\begin{aligned} G[h] &= \sum_{l=0}^{N-1} g[l]e^{\frac{-j2\pi hl}{N}}, \quad h = 0, 1, \cdots, N-1 \\ g[l] &= \frac{1}{N} \sum_{l=0}^{N-1} G[h]e^{\frac{j2\pi hl}{N}}, \quad l = 0, 1, \cdots, N-1 \end{aligned}$$

Generally, if $g_d[l]$ is nonzero only over a finite domain, $G[h]$ equals $G_d(w)$ at

equally spaced intervals of w:

$$G[h] = G_d(2\pi h/N), \quad h = 0, 1, \cdots, N-1$$

Increasing the DFT size N, e.g. by zero-padding, helps the DFT spectrum to become closer to the DTFT spectrum because more DTFT samples are obtained. This is known as the zero padding theorem, that zero padding in the time domain corresponds to ideal interpolation in the frequency domain.

Figure 4.4 illustrates the relationship between CTFT, DTFT and DFT.

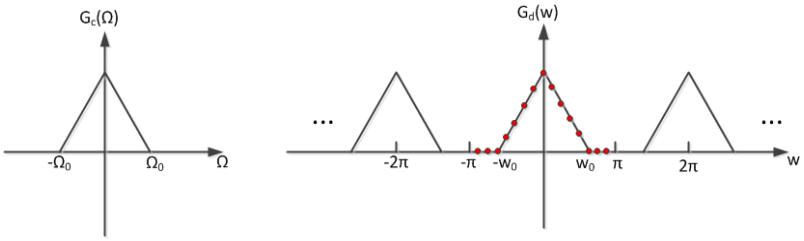


Figure 4.4: Plot of CTFT $G_c(\Omega)$ and DTFT $G_d(w)$ where $w_0 = \Omega_0 T$ with T as the sampling period, and the dots correspond to the DFT samples.

4.3.2 Formulation of Forward Projection for Parallel-beam CT

Suppose the sampling interval of the sinogram and the object is T_1 and T_2, respectively. CTFT and DTFT of the sinogram is $\hat{S}_c$ and $\hat{S}_d$ and that of the object is $\hat{F}_c$ and $\hat{F}_d$. According to FST, we have $\hat{S}_c(\rho) = \hat{F}_c(\rho\cos\theta, \rho\sin\theta)$, where ρ is the frequency index of the sinogram. According to (4.3), we obtain $T_1\hat{S}_d(\rho) = T_2\hat{F}_d(\frac{T_2}{T_1}\rho\cos\theta, \frac{T_2}{T_1}\rho\sin\theta)$. Assuming the image is of size $n \times n$, ROI is a circle of diameter n and the sinogram is obtained at q angular locations with p detectors, then the 1-D DFT of the full projection at each view corresponds to $\hat{S}_d$ at locations

$$\rho = \frac{2\pi h}{p}, h = 0, 1, \cdots, p-1$$

while in the Fourier domain of the image, it corresponds to $\hat{F}_d$ at locations:

$$(u_\theta = \frac{2\pi T_2 h\cos\theta}{T_1 p}, v_\theta = \frac{2\pi T_2 h\sin\theta}{T_1 p}), h = 0, 1, \cdots, p-1$$

which can hardly be mapped to an exact 2-D rectangular discrete system. Since the DFT of a $n \times n$ image defines a coarse Fourier space at $u, v = \frac{2\pi l}{n}, l = 0, 1, \cdots, n-1$, it is natural to pad the image with 0 to a fairly big size, as shown in Figure 4.5, so that a finer frequency resolution can be obtained and the coefficients at (u_θ, v_θ) can be put in this spectral structure with high accuracy.

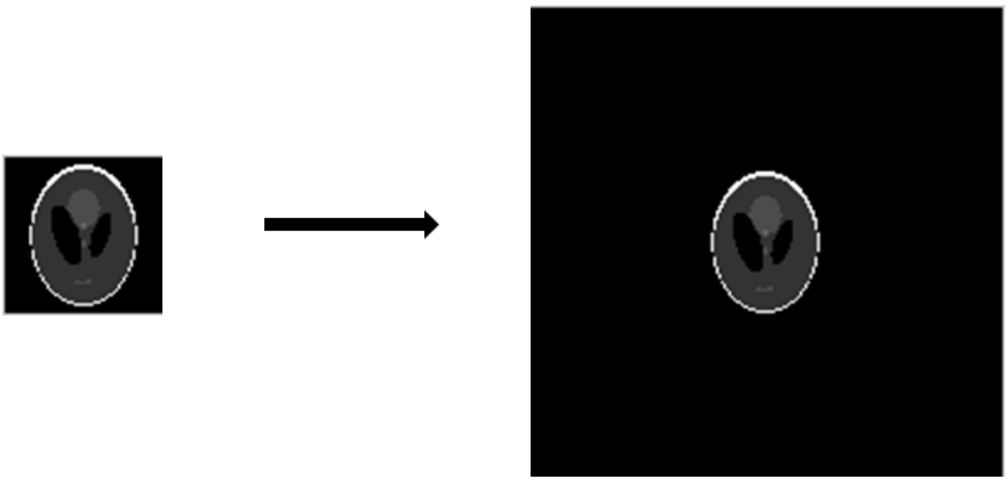


Figure 4.5: The left image is the original phantom, and the right image is after zero padding

Here we analyse the interpolation error. For simplicity, assume $p = n$, so that $T_1 = T_2$. Suppose the image is padded to $M \times M$ and $M = Zp$. Because the DTFT of the image is twice continuously differentiable, its linear approximation is obtained:

$$\hat{F}_d(u_\theta, v_\theta) \approx \hat{F}_d(\tilde{u}, \tilde{v}) + \frac{\partial \hat{F}_d}{\partial u}(\tilde{u}, \tilde{v})(\tilde{u} - u_\theta) + \frac{\partial \hat{F}_d}{\partial v}(\tilde{u}, \tilde{v})(\tilde{v} - v_\theta)$$

where $u_\theta = \frac{2\pi h \cos\theta}{p}, v_\theta = \frac{2\pi h \sin\theta}{p}, \tilde{u} = \frac{2\pi a}{pZ}, \tilde{v} = \frac{2\pi b}{pZ}, h = 0, 1, \cdots, p-1, a, b = 0, 1, \cdots, pZ-1$. Note that $h\cos\theta$ is approximated by $\frac{a}{Z}$ and $h\sin\theta$ by $\frac{b}{Z}$.

Since $\hat{F}_d(u, v) = \sum_x \sum_y f(x, y) e^{-j(ux+vy)}$, then

$$\begin{aligned} \hat{F}_d(u_\theta, v_\theta) \approx\ & \hat{F}_d(\tilde{u}, \tilde{v}) + \sum_x \sum_y f(x, y) x e^{-j(\tilde{u}x + \tilde{v}y + \frac{\pi}{2})}(\tilde{u} - u_\theta) \\ & + \sum_x \sum_y f(x, y) y e^{-j(\tilde{u}x + \tilde{v}y + \frac{\pi}{2})}(\tilde{v} - v_\theta). \end{aligned}$$

Therefore the interpolation error is:

$$\begin{aligned}
\text{Err} &= \sum_x \sum_y f(x,y) x e^{-j(\tilde{u}x+\tilde{v}y+\frac{\pi}{2})} \frac{2\pi}{p}(\frac{a}{Z} - h\cos\theta) \\
&\quad + \sum_x \sum_y f(x,y) y e^{-j(\tilde{u}x+\tilde{v}y+\frac{\pi}{2})} \frac{2\pi}{p}(\frac{b}{Z} - h\sin\theta) \\
&= \sum_x \sum_y f(x,y) e^{-j(\tilde{u}x+\tilde{v}y)} e^{-j\frac{\pi}{2}} \frac{2\pi}{p}[x(\frac{a}{Z} - h\cos\theta) + y(\frac{b}{Z} - h\sin\theta)]
\end{aligned}$$

Hence, $\text{Err} = \hat{F}_d(\tilde{u},\tilde{v})O(\frac{1}{Z})$, where O is the big O notation and means that the error decreases proportionally to the padding size.

At this stage we can break the forward projection down into the following steps:

1. Pad the $n \times n$ image to $M \times M$.
2. Apply 2-D DFT and the fine spectral structure of the image is defined at $u, v = \frac{2\pi l}{M}, l = 0, 1, \cdots, M-1$.
3. Pick the Fourier coefficients at (u_θ, v_θ) via nearest interpolation.
4. Apply 1-D inverse DFT to the Fourier coefficients of each view and obtain the sinogram of length p. Under-sample the projections at each angle according to the practical situation.

The matrix form of the forward projection can be expressed as:

$$\begin{pmatrix} \mathcal{P}_1 F_1^{-1} & & & \\ & \mathcal{P}_2 F_1^{-1} & & \\ & & \ddots & \\ & & & \mathcal{P}_q F_1^{-1} \end{pmatrix} \mathcal{P} F_2 \mathbf{x} = \mathbf{b} \tag{4.4}$$

where $\mathbf{x}$ is the column-wise vector of the object, and $\mathbf{b}$ is the projection vector. F_1^{-1} is the inverse 1-D Fourier matrix of size $p \times p$ and $\mathcal{P}_1, \mathcal{P}_2, \cdots, \mathcal{P}_q$ denote the detector sampling at each views, i.e. to select partial rows from each F_1^{-1}. It is simply a subset of $p_j (j = 1, 2, \cdots, q)$ rows of the $p \times p$ identity matrix, and p_j is the detector number at angular position θ_j. $\mathcal{P}$ is to pick the coefficients at (u_θ, v_θ), which is also a subset of $p \times q$ rows of the identity matrix of size $M \times M$. F_2 denotes the 2-D Fourier transform of the $M \times M$ image, sampled

at column due to the zero-padding, so its size is $M^2 \times n^2$. The accuracy of the model in (4.4) is verified in the simulation section, where it is shown to produce comparable sinograms to the Matlab built-in model and the linear intersection model.

4.4 Sampling Schemes Reflected in the Model

4.4.1 Rebinning from Fan-beam CT to Parallel-beam CT

The projection model (4.4) reflects well the under-sampling schemes for parallel-beam CT, with the detector r sampling corresponding to $\mathcal{P}_1, \mathcal{P}_2, \cdots, \mathcal{P}_q$ and the angle θ sampling corresponding to $\mathcal{P}$.

Since our focus here is to study the effect of different down-sampling schemes on compressed sensing-based reconstruction in fan-beam CT, it is necessary to rebin the fan-beam CT projections to parallel-beam CT. The data acquisition geometry [38] is illustrated in Figure 4.6. It shows that the X-ray source travels along a circular trajectory, denoted by $\mathbf{a}(\beta) = (R\cos\beta, R\sin\beta)^T$, where $\beta \in [0, 2\pi)$ is the rotation angle, R is the distance between the source and the rotation axis, and the superscript T stands for the transpose operator. The curved detector array lies on a circle of radius D centred at the source. Any ray in the fan beam is specified by an angle $\gamma \in [-\gamma_m, \gamma_m]$ that increases in the clockwise direction. The definition of γ_m is as follows. Assuming the object is compactly supported in a centred circular field of view (FOV), the value of $|\gamma|$ for each of the two rays tangent to the FOV is denoted by γ_m.

The relation between the independent variables of the fan-beam projections and parallel projections is

$$r = R\sin\gamma \quad \text{and} \quad \theta = \beta + \gamma, \tag{4.5}$$

and it is plotted in Figure 4.7 when β covers over 180°. Because of the periodic property of parallel projections, the regions labeled A are equivalent, with the upper part for $\theta > 180°, r > 0$ and the lower part for $\theta < 0°, r < 0$. On the other hand, the regions marked B are areas in the parallel-projection space where no measurements are taken in fan-beam CT. Hence, in order to cover the full parallel

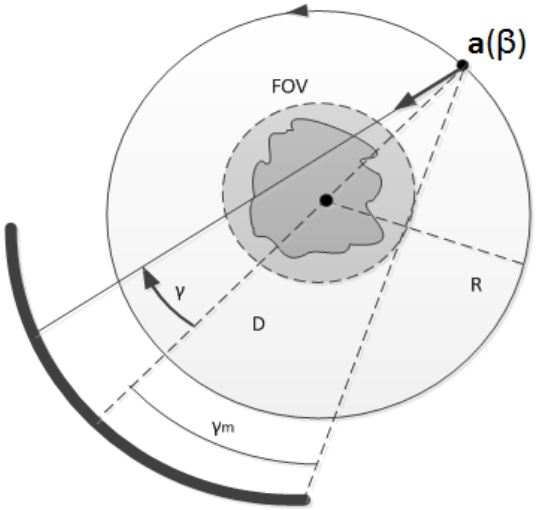


Figure 4.6: Fan-beam acquisition geometry with a curved detector

projections , $\beta \in [-\gamma_m, 180° + \gamma_m)$.

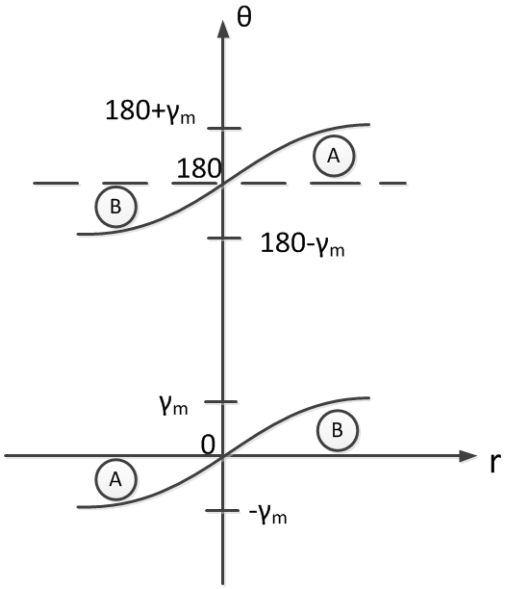


Figure 4.7: Fan-beam projections collected at $\beta \in [0° \quad 180°]$ give estimates of the parallel projections between the curved lines.

4.4.2 Models under Different Sampling Schemes

As the projection model for parallel-beam CT is formulated as

$$\begin{pmatrix} \mathcal{P}_1 F_1^{-1} & & & \\ & \mathcal{P}_2 F_1^{-1} & & \\ & & \ddots & \\ & & & \mathcal{P}_q F_1^{-1} \end{pmatrix} \mathcal{P} F_2 \mathbf{x} = \mathbf{b},$$

suppose $A_1 = \begin{pmatrix} \mathcal{P}_1 F_1^{-1} & & & \\ & \mathcal{P}_2 F_1^{-1} & & \\ & & \ddots & \\ & & & \mathcal{P}_q F_1^{-1} \end{pmatrix}$ and $A_2 = \mathcal{P} F_2$, first we investigate what A_1, A_2 are like under the full sampling of fan-beam CT, i.e. when the scanning comes with $2n$ samples in both the view and bin direction. Suppose the step of β and γ is denoted by Δ_β and Δ_γ, we have

$$\begin{aligned} &\because \quad N_{\text{views}} = N_{\text{bins}} = 2n \\ &\therefore \quad \Delta_\gamma = 2\gamma_m/(2n-1), \Delta_\beta = (180 + 2\gamma_m)/(2n-1) \\ &\therefore \quad \Delta_\beta = (90/\gamma_m + 1)\Delta_\gamma \end{aligned}$$

If the X-ray source is placed at $R = 2n$ away from the centre, which is a popular and reasonable setting, we obtain $\gamma_m \approx 18^\circ$, then $\Delta_\beta = 6\Delta_\gamma$, as illustrated in Figure 4.8. The grid intersections denote the measured fan projections. For the ease of the analysis of the sampling schemes, it is assumed that the rebinned projections are obtained at θ with the increment Δ_β. The desired parallel projections are shown in blue in Figure 4.8 and each line stands for the same θ, from 0° at the bottom to 180° at the top. Hence, the $2n \times 2n$ fan projections are converted to parallel projections on $\frac{5}{3}n + \frac{1}{6}$ views and $2n$ detectors at each view.

Based on the rebinning analysis above, for the full scanning of fan-beam CT, A_1 is the block diagonal matrix composed of $\frac{5}{3}n + \frac{1}{6}$ 1D inverse Fourier transform matrices, each of size $2n \times 2n$. This makes the size of $A_1 \quad (2n(\frac{5}{3}n + \frac{1}{6}))^2$. Since there is no under-sampling, $\mathcal{P}_1, \mathcal{P}_2, \cdots, \mathcal{P}_q$ are all identity matrices. The function of A_2 is to pick the Fourier coefficients lying on $\frac{5}{3}n + \frac{1}{6}$ angles of the zero-padded $n \times n$ images, therefore A_2 is of the size $2n(\frac{5}{3}n + \frac{1}{6}) \times n^2$. The 2-D pattern is shown in Figure 4.9. It is a full-column-rank matrix.

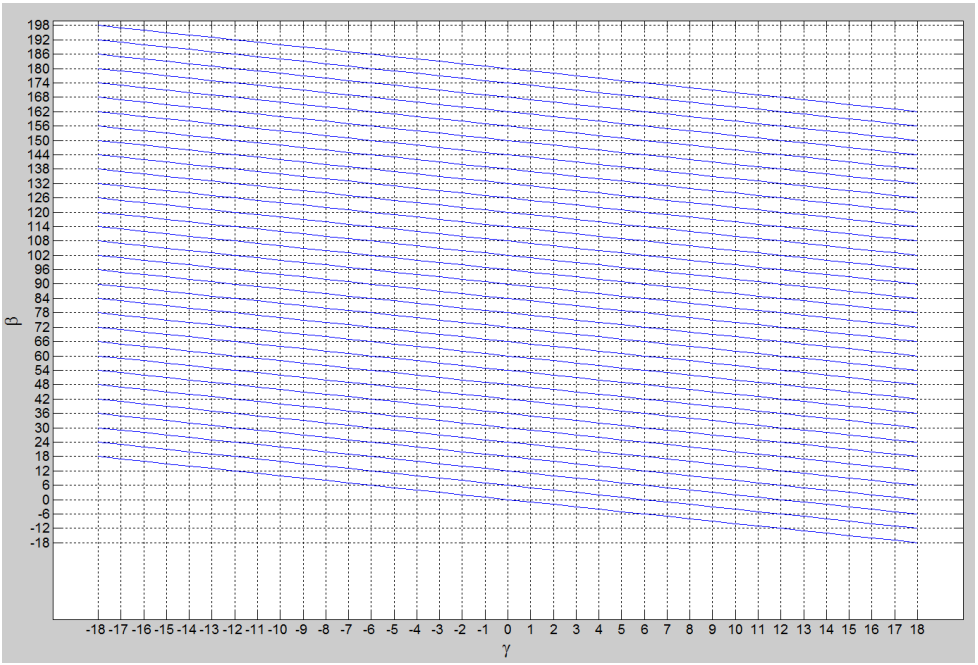


Figure 4.8: Relationship between parallel and fan projections when $\Delta_\beta = 6\Delta_\gamma$

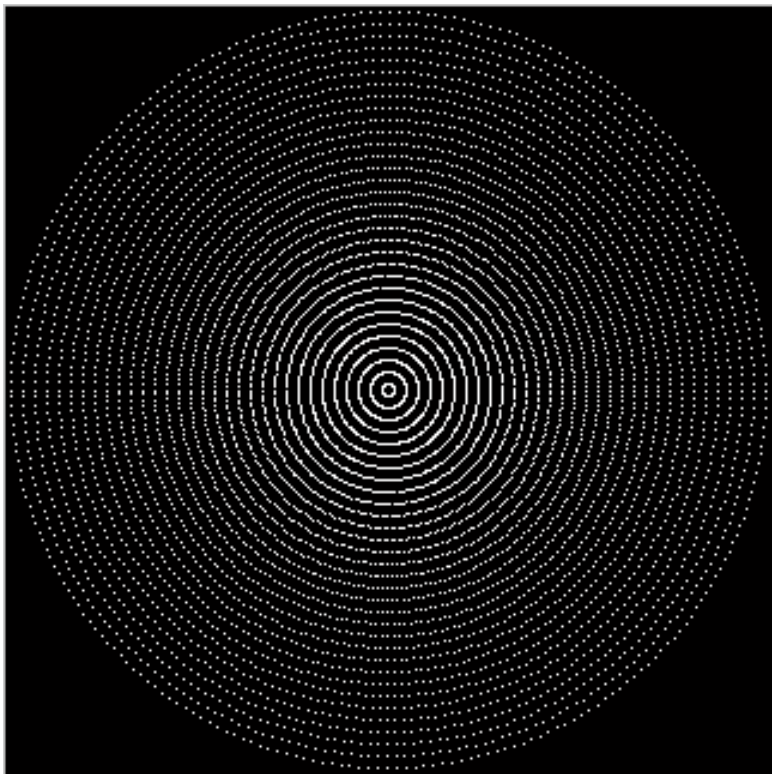
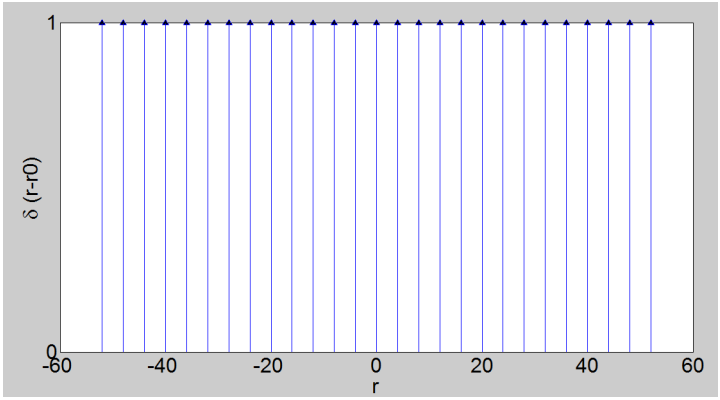


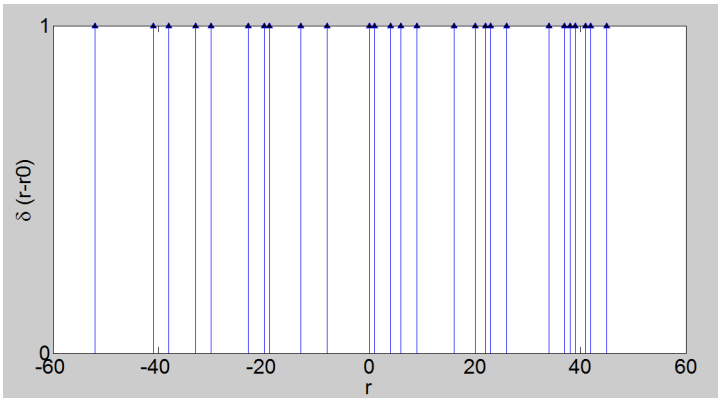
Figure 4.9: Sampling pattern of A_2

Next we consider how the detector and angle sampling in fan-beam CT affects the model. As (4.5) reveals that r is an injective function of γ, the sampling pattern

reflects directly on $\mathcal{P}_1, \mathcal{P}_2, \cdots, \mathcal{P}_q$ in detector sampling. For regular-detector and random-detector schemes, $\mathcal{P}_1, \mathcal{P}_2, \cdots, \mathcal{P}_q$ present a uniform distribution as in Figure 4.10 (a) and a random distribution as in Figure 4.10 (b), respectively.



(a) *regular-detector scheme*

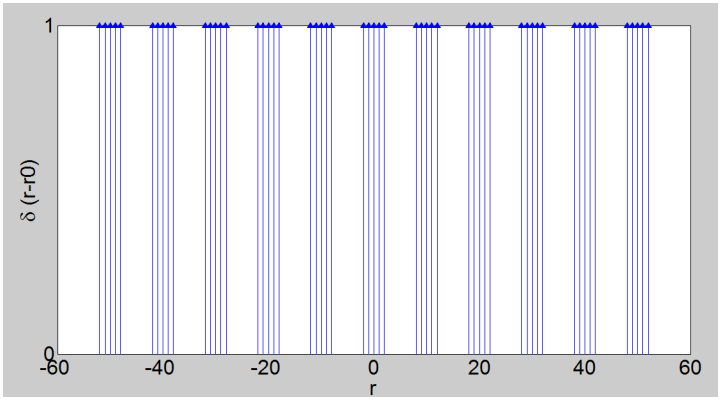


(b) *random-detector scheme*

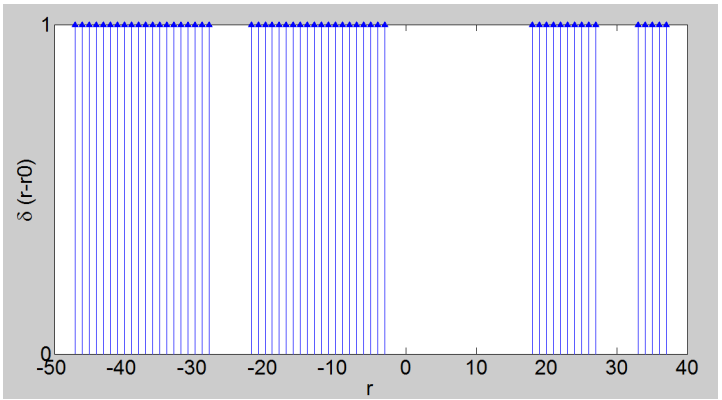
Figure 4.10: Different detector-sampling schemes

On the other hand, as the projections at the same β affect 5-6 projections at the same θ, angle sampling would cause cluster distribution on each view, shown in Figure 4.11. In the golden-angle scheme, every θ is sampled by a similar ratio; while the random-angle scheme will create heavy sampling at some θ and light sampling at other θ.

In summary, the four schemes only affect A_1 and the A_2 maintains the full-column-rank status as in the full sampling case. How A_1 behaves is listed in Table 4.1.



(a) *golden-angle scheme*



(b) *random-angle scheme*

Figure 4.11: Different angle-sampling schemes

Table 4.1: Effects of sampling patterns on the model

sampling schemes	pattern of $\mathcal{P}_1, \mathcal{P}_2, \cdots, \mathcal{P}_q$
regular-detector scheme	same uniform sampling at every view
random-detector scheme	different random sampling at every view
golden-angle scheme	different uniform cluster sampling at every view
random-angle scheme	different random cluster sampling at every view

4.5 Simulation

The sinogram of several different images is obtained by three different models: the Matlab built-in model (using pixel-driven methods), linear intersection model (using ray-driven methods) and the FST-based model formulated in this chapter. In the first model, the algorithm first divides pixels in the image into four

subpixels and projects each subpixel separately, as shown in Figure 4.12. Each subpixel's contribution is proportionally split into the two nearest bins, according to the distance between the projected location and the bin centres. If the subpixel projection hits the centre point of a bin, the bin on the axes gains the full value of the subpixel, or one-fourth the value of the pixel. If the subpixel projection hits the border between two bins, the subpixel value is split evenly between the bins. In the linear intersection model, the CT system matrix is composed of the intersection length of the X-ray and the image lattices, as described in Section 2.5.1. Four images are selected in Figure 4.13 and Figure 4.14 to show that the FST-based model can obtain comparable results with the current two most popular models.

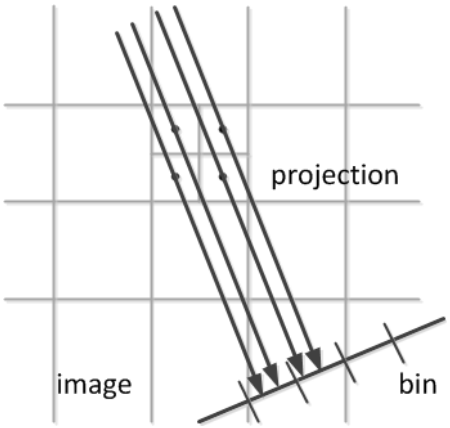
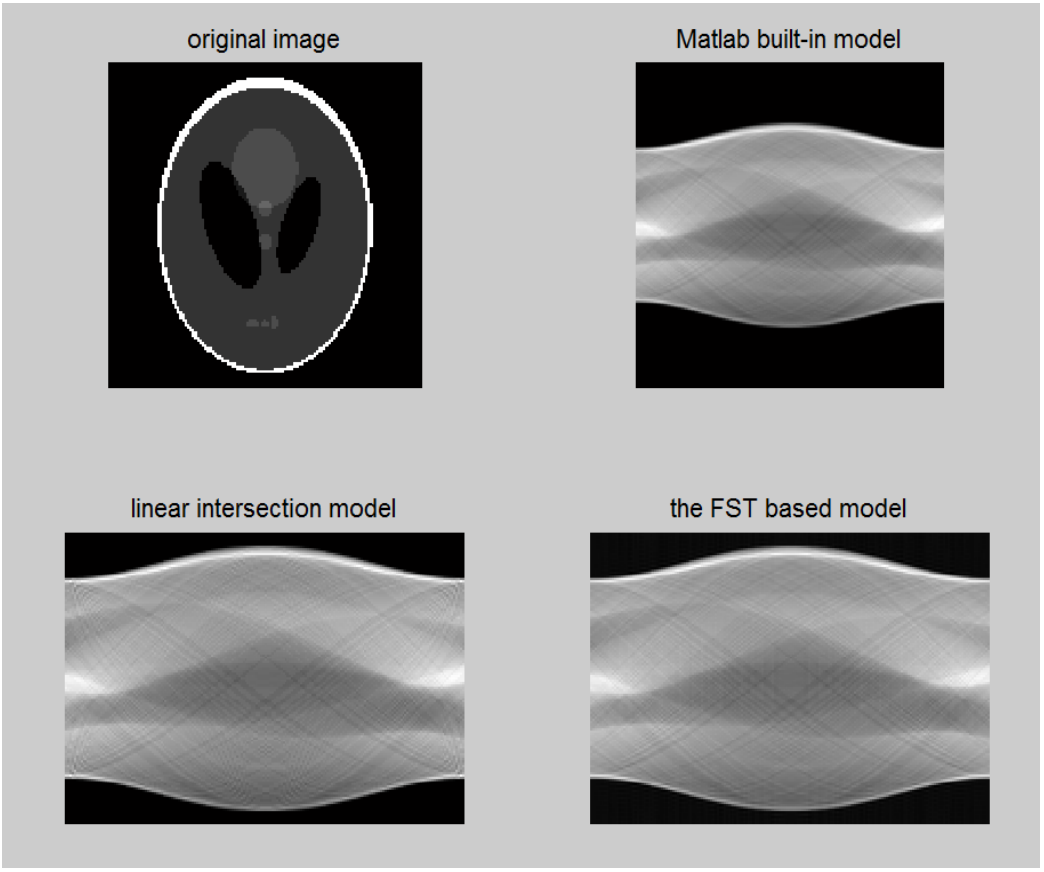


Figure 4.12: Matlab built-in model

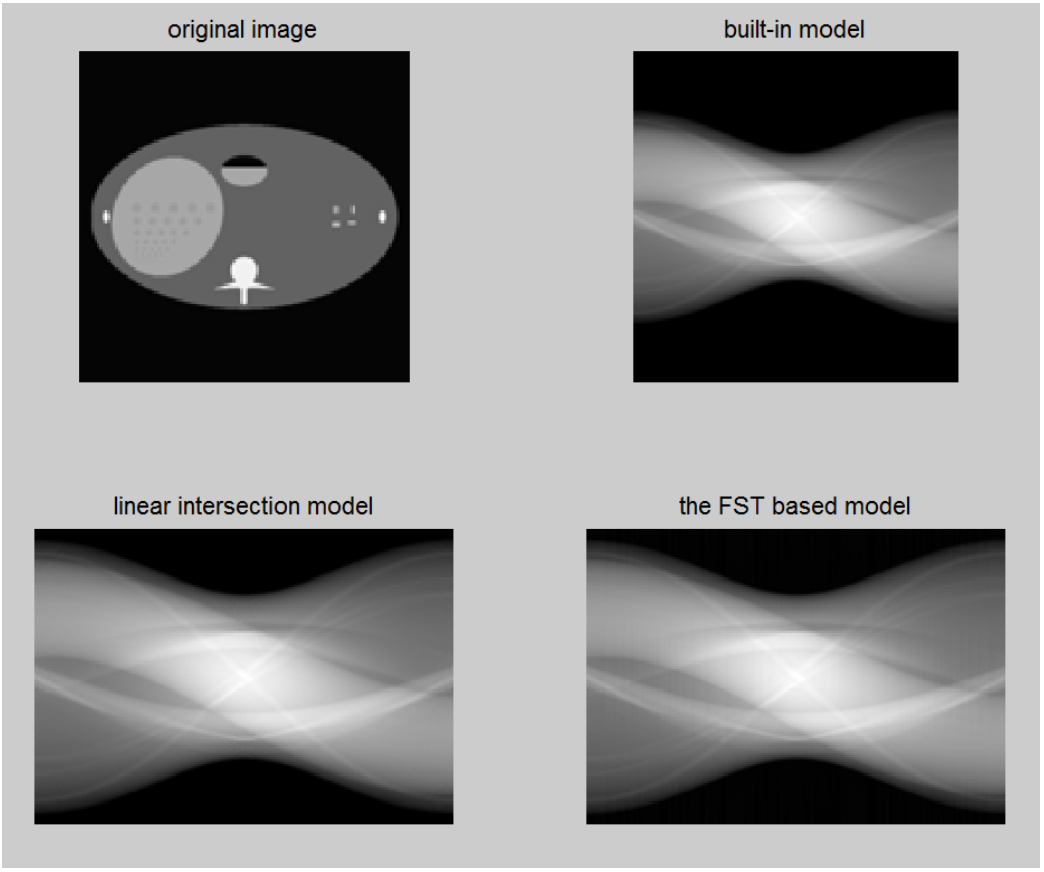
4.6 Conclusion

In this chapter, we have analysed the sampling schemes in CT. First, the full scanning fan-beam CT was illustrated, followed by the introduction of the sparse-view and sparse-detector CT, leading to four specific sampling schemes: golden-angle and random-angle, and regular-detector and random-detector. Next we built the projection model of parallel-beam CT based on the preliminary knowledge of FST and the relation between different Fourier transforms, and error analysis of the model showed that the error becomes negligible when the padding size is sufficiently large. In the following section, we bridged the gap between the model and fan-beam CT by revealing the rebinning relationship between fan-beam CT and parallel-beam CT, and depicted in detail what the model is like for full-scanning,

sparse-view and sparse-detector fan-beam CT, respectively. Finally, a simple simulation was run to validate the accuracy of the model compared to the Matlab built-in model and the linear intersection model.

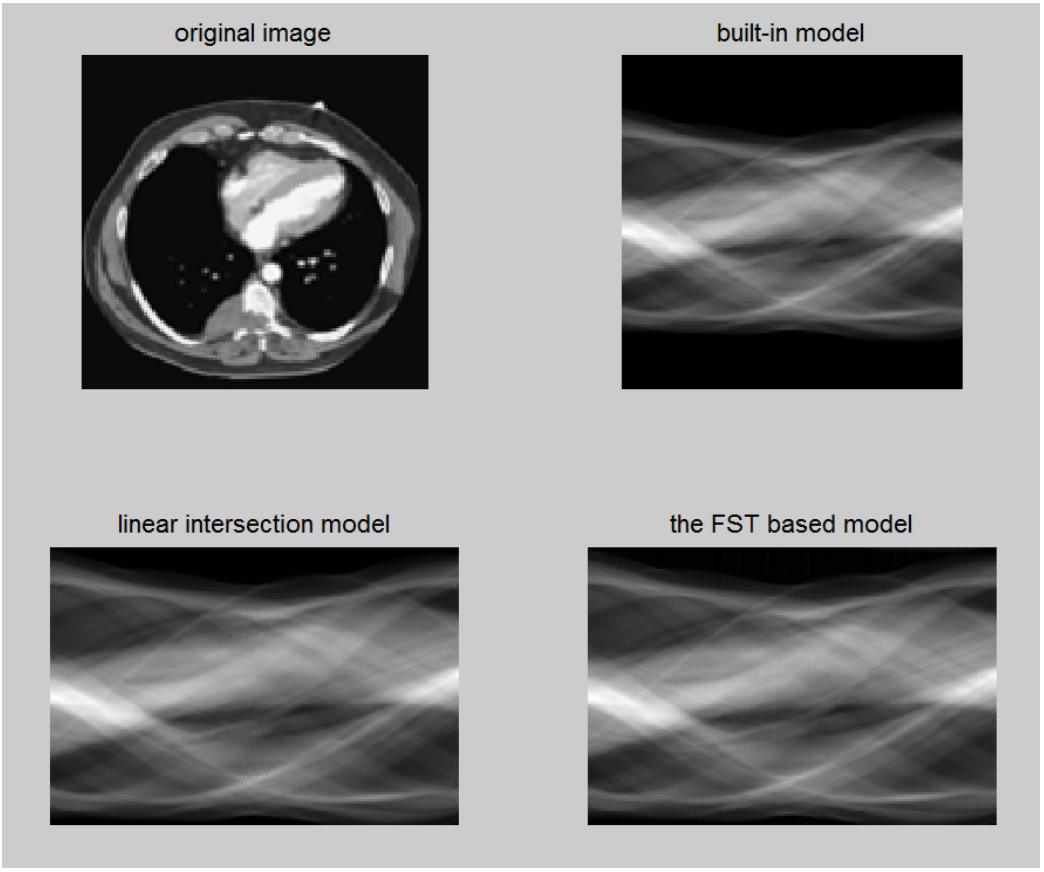


(a) *phantom1*

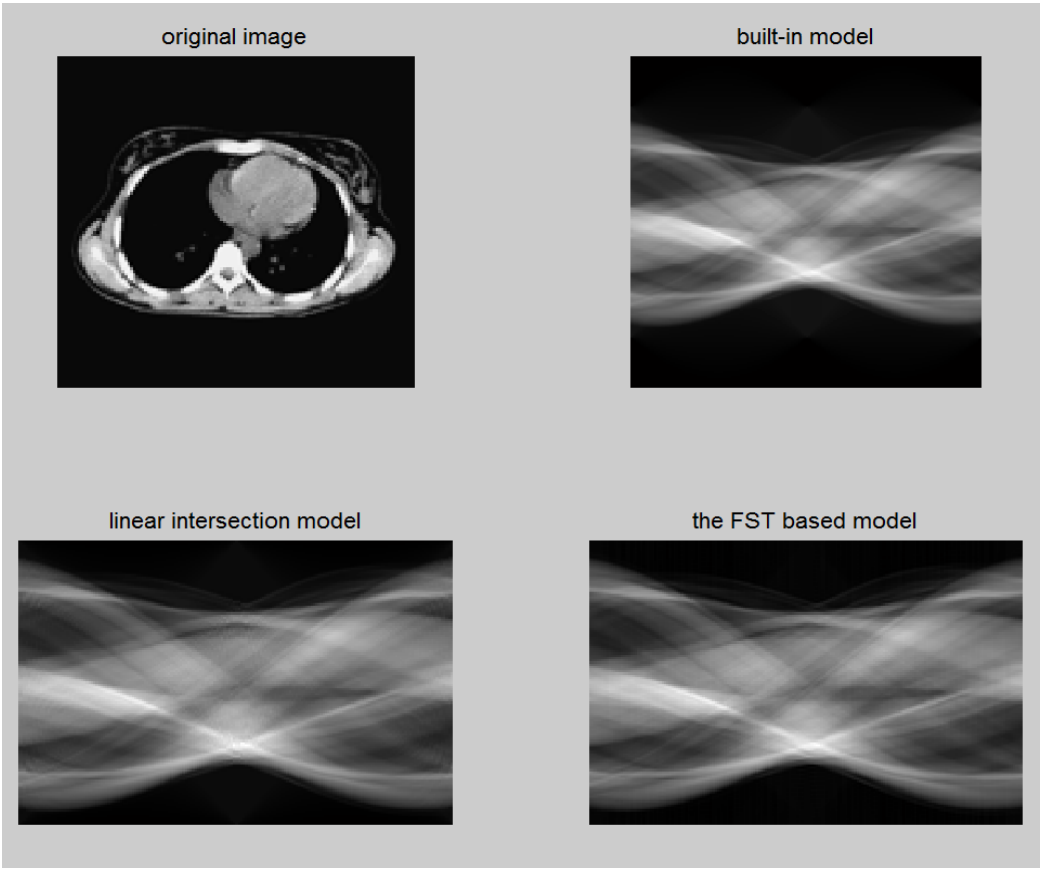


(b) *phantom2*

Figure 4.13: Sinogram of phantoms obtained by different models



(a) *CT1*



(b) *CT2*

Figure 4.14: Sinogram of CT images obtained by different models

Chapter 5

Analysis of Compressed Sensing-Based CT Reconstruction with Low Radiation

5.1 Introduction

As a popular and useful medical tool, X-ray computed tomography (CT) plays an important role in diagnosis and thus benefits patients greatly. In general, the image quality of CT is proportional to the radiation dose. However, it is known that X-ray radiation leads to ionization of body cells and increased radiation dose raises the risk of cancer. Therefore, reduction of the X-ray radiation dose to patients while preserving the imaging quality has been a significant and challenging problem in CT development.

Conventional CT acquires projection data over a number of view angles and uses filtered back-projection (FBP) [168] for reconstruction. The number of X-ray projections and data acquisitions for satisfactory image reconstruction is determined by the well known Shannon-Nyquist sampling theorem. Recently, iterative reconstruction techniques (IRT) [142, 158, 169] have been shown to be capable of reconstructing the image using far few data samples. Specifically, compressed

sensing (CS) technology has been demonstrated to be a powerful technology for recovering signals from incomplete measurements through optimization methods and has been applied to many applications. A number of CS-based techniques have been applied to CT reconstruction problems subject to under-sampled or noisy data [29, 70, 139, 140]. The key ideas are that CT images can be sparsely represented by some linear transformation and penalizing the norm of the image in the sparsified domain can enable recovery of the unknown image from highly under-sampled data. The application of CS provides great benefits in radiation reduction, and therefore has become one of the central topics in medical imaging [61, 68].

The existing studies of CT and CS can be roughly divided into two classes. One class is devoted to reconstruction accuracy through different optimization conditions and algorithms [14, 61, 68, 88, 89], the other is devoted to reconstruction speed and data storage [70, 110, 116] as CT, especially 3-D CT, requires a large system matrix. What is seldom discussed, as mentioned in [73], is that the theoretical results from CS do not extend to the CT setting. There is a fundamental lack of understanding about which type of under-sampling is favourable for post CS reconstruction. The reason is that CS requires the sampling matrix in the sparsifying domain to meet the restricted isometry property (RIP), which cannot be tested by any known polynomial-time algorithm. In other words, it is the key challenge to achieve incoherence in designing CT data acquisition methods for successful CS reconstruction [76]. Sidky et al. [138] try to tackle the restricted isometry properties of the Radon transform. Firstly they solve the problem that isometry constants are not invariant to scalings of the system matrix, followed by the design of a numerical strategy for determining a lower bound on the RIP constant of a CT matrix for a 512^2-pixel image. Because Radon transform has the greatest difficulty in distinguishing neighbouring pixels, the search of isometry constants involves only neighbouring pixels. The proposed strategy on one hand requires intensive computation, and on the other hand is not very practical, as the lower bound is already approaching 1 at sparsity of 100. Jørgensen, Sidky and Pan [73] develop quantitative notions for under-sampling based on the condition number of the system matrix, which addresses only invertibility and stability. Lustig, Donoho and Pauly [94] use the maximum of the sidelobe-to-peak ratio of the point spread function (PSF) to measure the incoherence. In this chapter, we select the system matrix class for a 2-D circular fan-beam geometry using a square-pixel array and investigate the property of the under-sampled system

matrix in light of RIP with two tools: PSF and the Fourier slice theorem (FST). The main contribution here is to provide understanding of CS-based CT image reconstruction and guidelines for the design of low radiation CT scanning and reconstruction.

The rest of the chapter is organized as follows. Section 5.2 gives a brief introduction to fan-beam CT data acquisition, and the CS-based framework for CT reconstruction. Different sampling schemes and the analysis of their impacts on reconstruction are presented in Section 5.3. The specified optimization algorithm TVAL3 is described in Section 5.4. In Section 5.5, the experimental results show consistency with the theoretical analysis and the index proposed is more sensitive than that in [94]. Conclusions are drawn in Section 5.6.

5.2 Background

5.2.1 Fan-beam Computed Tomography

The main part of a typical CT scanner is a doughnut-shaped gantry that consists of a set of X-ray sources and detectors on opposite sides. The sources emit X-rays that are attenuated when passing through the object and then detected at the detectors, from which an intensity map of the transmission coefficients at various points in the object can be reconstructed. There are different configurations of X-ray sources and detectors, from the early generation of parallel-beam scan to the currently popular cone-beam spiral scan. In this chapter, we take fan-beam projections for simplicity and the results can be extended to other configurations. The basic configuration of fan-beam CT is as follows: a point source of radiation emanates a fan-shaped beam, and on the other side of the object a bank of detectors acquires all the measurements in one fan simultaneously. The source and the entire bank of detectors are rotated to generate the desired number of fan projections. The data acquisition geometry is the same as in the previous chapter, illustrated in Figure 4.6.

5.2.2 Regularized Optimization for CS-based CT Reconstruction

Compressed sensing [130] is an emerging technology which can achieve perfect signal recovery from much fewer measurements than usually required by the Shannon-Nyquist criterion. It exploits the sparseness or compressibility of signals in a predefined basis/frame and its required number of measurements is proportional to the number of non-zero elements in the sparse representation of signals [24]. The CS-based CT reconstruction problem can be formulated as:

$$min \quad ||\varphi \mathbf{x}||_1 \quad s.t. \quad A\mathbf{x} = \mathbf{b} \tag{5.1}$$

where φ is the sparsifying transform, $\mathbf{x}$ is the column-wise vector of the object, $\mathbf{b}$ is the projection vector and A is the system matrix based on the linear intersection model. When the observation is noisy, the object can be reconstructed through:

$$min \quad ||\varphi \mathbf{x}||_1 + \frac{\tau}{2}||A\mathbf{x} - \mathbf{b}||_2^2 \tag{5.2}$$

where $\tau > 0$ is the penalty parameter. The first part in the objective function is the regularization term and the second is often referred to as the fidelity term to ensure data consistency. φ is dependent on image property, and currently total variation (TV) regularization [132] is a popular choice. TV regularization succeeds when the gradient of the underlying image is sparse, i.e. the image possesses piecewise constant property. However, the properties of non-differentiability and non-linearity of TV functions make the problem solving more challenging computationally than ℓ_1 minimization models. Here the optimization algorithm TVAL3 [91] is adopted, which is favourable in terms of reconstruction quality and speed. With the input of the sampling matrix A and the measurements $\mathbf{b}$, it approximately minimizes the augmented Lagrangian function by an alternating direction scheme and updates multipliers until the reconstruction error is sufficiently small. The details are described in Section 5.4.

5.3 Impacts of Sampling Schemes on CS-based Reconstruction

Specific sampling schemes for the purpose of radiation reduction have been discussed in Chapter 4. In this section, their impact on CS-based reconstruction is investigated with reference to the optimal low-radiation CT system design. The study is carried out from two aspects. One is based on the point spread function (PSF) of the system matrix, and the other is to analyze the sampling patterns of Fourier matrix, which relates closely to the system matrix presented in Chapter 4.

5.3.1 Index Based on PSF

RIP claims that if $\delta_{2k} \leq \sqrt{2}-1$, we can find a unique k-sparse solution for $\mathbf{b} = A\mathbf{x}$. There is currently no known polynomial-time algorithm to test whether a given matrix satisfies the RIP. Hence, an index is developed in this section as the indication of the isometry constant.

Theorem 5.1. *The isometry constant δ_k of $A \in \mathbb{C}^{m\times N}$ with normalized columns possesses the following property:*

$$\delta_k = \max_{\mathbf{K}\subset[N],|\mathbf{K}|\leq k} ||A_{\mathbf{K}}^* A_{\mathbf{K}} - I||_2 \leq ||A^*A - I||_F$$

where $[N] := 1, 2, \cdots, N$ and $|\mathbf{K}|$ is the cardinality of $\mathbf{K}$, $A_{\mathbf{K}} = (a_j)_{j\in\mathbf{K}}$, i.e. it is the column sub-matrix of A consisting of the columns indexed by $\mathbf{K}$.

Proof. The definition of the isometry constant is equivalent to [125]

$$|||A\mathbf{x}||_2^2 - ||\mathbf{x}||_2^2| \leq \delta_k ||\mathbf{x}||_2^2,$$

$$\forall \mathbf{K} \subset [N], |\mathbf{K}| \leq k, \forall \mathbf{x} \in \mathbb{C}^N, \text{supp} \quad \mathbf{x} \subset \mathbf{K}$$

The left hand side equals $| < (A^*A - I)\mathbf{x}, \mathbf{x} > |$. Taking the supremum over all $\mathbf{x} \in T_k = \{\mathbf{x} \in \mathbb{C}^N, ||\mathbf{x}||_2 = 1, ||\mathbf{x}||_0 \leq k\}$ shows:

$$\delta_k \quad = \quad \sup_{X\in T_k} | < (A^*A - I)\mathbf{x}, \mathbf{x} > |$$

$$
\begin{aligned}
&= \max_{\mathbf{K}\subset[N],|\mathbf{K}|\leq k} ||A_{\mathbf{K}}^* A_{\mathbf{K}} - I||_2 \\
&\leq \max_{\mathbf{K}\subset[N],|\mathbf{K}|\leq k} ||A_{\mathbf{K}}^* A_{\mathbf{K}} - I||_F \\
&\leq ||A^* A - I||_F
\end{aligned}
$$

□

According to the theorem, instead of calculating the isometry constant directly, we can measure the Frobenius norm of $A^*A - I$. A^*A is recognized as the point spread function (PSF) in Donoho's work [45]:

$$\mathrm{PSF}(i;j) = e_j^* A^* A e_i, \tag{5.3}$$

where e_i and e_j is the ith and jth vector of natural basis, respectively. PSF is a natural tool for measuring incoherence. It measures the contribution of a unit-intensity pixel at the ith position to a pixel at the jth position [130]. With Nyquist sampling there is no interference between pixels and $\mathrm{PSF}(i;j)_{i\neq j} = 0$. Under-sampling causes pixels to interfere and assigns nonzero values to $\mathrm{PSF}(i;j)_{i\neq j}$. A nonzero value of PSF at $(i;j)$ means that linear reconstruction of pixel i suffers interference by a unit impulse at pixel $j \neq i$. Since CS requires the aliasing artifacts due to under-sampling to be incoherent, we expect $\mathrm{PSF}(i;j)_{i\neq j}$ to be noise-like. From the energy viewpoint, the PSF measures the tendency of zero-filled linear reconstruction to leak energy from the true underlying source pixel to other pixels. This energy is shown as blurring or aliasing artifacts in the reconstructed image. In designing the sampling matrix, such energy leakage is expected to spread quasi-uniformly across the image.

To determine the impacts of sampling schemes on reconstruction quality, the Frobenius norm of $A^*A - I$, noted as Υ , is adopted as the quantified index. It is shown as follows in terms of PSF with normalized-column A:

$$\Upsilon = ||\mathrm{PSF} - I||_F \tag{5.4}$$

Satisfactory reconstruction can be expected when Υ is small.

5.3.2 Analysis Based on FST

It was shown in Chapter 4 that Fourier slice theorem (FST) perfectly connects the CT system matrix and the Fourier matrix. Hence, it is feasible to analyze the sparse sampling effect based on the RIP of the Fourier transform, which is well defined. Recall that the projection model is formulated as

$$\begin{pmatrix} \mathcal{P}_1 F_1^{-1} & & & \\ & \mathcal{P}_2 F_1^{-1} & & \\ & & \ddots & \\ & & & \mathcal{P}_q F_1^{-1} \end{pmatrix} \mathcal{P} F_2 \mathbf{x} = \mathbf{b},$$

and we set $A_1 = \begin{pmatrix} \mathcal{P}_1 F_1^{-1} & & & \\ & \mathcal{P}_2 F_1^{-1} & & \\ & & \ddots & \\ & & & \mathcal{P}_q F_1^{-1} \end{pmatrix}$, $A_2 = \mathcal{P} F_2$. We listed in Table 4.1 the effects of sampling patterns on A_1 and A_2, and now we investigate how it affects the reconstruction.

The four sampling patterns all conform to the case where A_1 is sampled while A_2 maintains the full-sampling status, which results in a synthesis model with the dictionary A_2:

$$\hat{\mathbf{x}}_s = \min_{\mathbf{x}} \quad \lambda ||\mathbf{x}||_1 + \frac{1}{2} ||A_1 A_2 \mathbf{x} - \mathbf{b}||_2^2 \tag{5.5}$$

The model (5.5) is nearly equivalent [47] to the analysis model with a full row-rank dictionary $A_2^+ = (A_2^T A_2)^{-1} A_2^T$:

$$\hat{\mathbf{x}}_a = A_2^+ (\min_{\mathbf{y}} \quad \lambda ||A_2^+ \mathbf{y}||_1 + \frac{1}{2} ||A_1 \mathbf{y} - \mathbf{b}||_2^2) \tag{5.6}$$

The proof is in Section 5.3.3.

The initial CS theory holds for signals that are sparse in the standard coordinate basis or sparse with respect to some orthonormal basis. However, in the above analysis model (5.6), A_2^+ is obviously with correlated columns. Here we need some improved results from the authors of [20], who state that CS is also viable in this

context. It is proposed to reconstruct from noisy measurements $\mathbf{b} = A\mathbf{x} + \mathbf{z}$ via:

$$\hat{\mathbf{x}} = \min_{\mathbf{x}} ||\psi^*\mathbf{x}||_1 \quad \text{s.t.} \quad ||A\mathbf{x} - \mathbf{b}||_2 \leq \epsilon \tag{5.7}$$

where ψ is an arbitrary tight frame, ϵ is a likely upper bound on the noise power $||\mathbf{z}||_2^2$.

Definition 5.2 (ψ-RIP)**.** *Let $\sum_k$ be the union of all subspaces spanned by all subsets of k columns of ψ, then the measurement matrix A obeys the restricted isometry property adapted to ψ (abbreviated ψ-RIP) with constant δ_k if*

$$(1 - \delta_k)||\mathbf{v}||_2^2 \leq ||A\mathbf{v}||_2^2 \leq (1 + \delta_k)||\mathbf{v}||_2^2$$

holds for all $\mathbf{v} \in \sum_k$.

Theorem 5.3. *Let A be a measurement matrix satisfying ψ-RIP with $\delta_{2k} < 0.08$. Then the solution to* (5.7) *obeys*

$$||\hat{\mathbf{x}} - \mathbf{x}|| \leq C_0\epsilon + C_1 \frac{||\psi^*\mathbf{x} - (\psi^*\mathbf{x})_k||_1}{\sqrt{k}}$$

where the constants C_0 and C_1 may only depend on δ_{2k}, $(\psi^*\mathbf{x})_k$ is the vector consisting of the largest k entries of $\psi^*\mathbf{x}$ in magnitude.

The assumption that ψ is a tight frame is not necessary and is proposed to simplify the analysis. Theorem 5.3 shows that ℓ_1-analysis is accurate when the coefficients of $\psi^*\mathbf{x}$ are sparse or decay rapidly. To return to our problem (5.6), we can now analyze it with standard CS knowledge, which states that random partial Fourier matrix can gain satisfactory reconstruction through CS.

For detector sampling, the sampling pattern reflects directly on $\mathcal{P}_1, \mathcal{P}_2, \cdots, \mathcal{P}_q$. It is preferable that $\mathcal{P}_1, \mathcal{P}_2, \cdots, \mathcal{P}_q$ present random distribution rather than regular distribution.

Angle sampling causes cluster distribution at each view of the parallel projections because the fan projections at the same β correspond to a small cluster at the same θ. That β is randomly distributed will make heavy sampling at some θ and light sampling at other θ, while every θ is sampled by a similar ratio at the even β case. In the Theorem III.1 proposed in [117], it is pointed out that the possibility of the distinct block diagonal (DBD) matrices satisfying RIP is proportional to

$\Gamma_2(\mathbf{Y}, \mathbf{Q})$. Here $\mathbf{Y}$ is the signal being measured that can be divided into q blocks $[\mathbf{Y}_1 \quad \mathbf{Y}_2 \quad \cdots \quad \mathbf{Y}_q]$, and the definition of $\Gamma_2(\mathbf{Y}, \mathbf{Q})$ is as follows:

$$\Gamma_2(\mathbf{Y}, \mathbf{Q}) := \frac{(\sum_{j=1}^{q} ||\mathbf{Y}_j||_2^2)^2}{\sum_{j=1}^{q} \frac{||\mathbf{Y}_j||_2^4}{Q_j}}$$

where $\mathbf{Q} = [Q_1, Q_2, \cdots, Q_q]$ is the measurement allocation. The bigger $\Gamma_2(\mathbf{Y}, \mathbf{Q})$, the more likely the DBD matrices will satisfy RIP. The theorem is for DBD matrices composed of Gaussian matrix, but the influence of the signal property should be the same for DBD matrices composed of Fourier matrix. In our case, the $[\mathbf{Y}_1 \quad \mathbf{Y}_2 \quad \cdots \quad \mathbf{Y}_q]$ are the coefficients on the lines of q angles at the Fourier domain of the image, hence the similar energy. i.e. similar $||\mathbf{Y}_j||_2^2$. To obtain bigger $\Gamma_2(\mathbf{Y}, \mathbf{Q})$, smaller $\sum_j \frac{1}{Q_j}$ is preferred. It is easy to know that $\sum_j \frac{1}{Q_j}$ achieves the minimum value when all Q_j is equal based on the premise that $\sum_j Q_j$ stays the same. Hence, the random β case is expected to have lower reconstruction quality due to the higher coherence at the heavy sampling point.

Based on the theoretical analysis of FST and the RIP of Fourier matrix, among the four sampling schemes, random-detector and regular-angle gain more satisfactory reconstructions than regular-detector and random-angle, respectively.

5.3.3 Proof of Equivalence Between the Analysis and the Synthesis Model

The synthesis model (5.5)

$$\hat{\mathbf{x}}_s = \min_{\mathbf{x}} \quad \lambda||\mathbf{x}||_1 + \frac{1}{2}||A_1 A_2 \mathbf{x} - \mathbf{b}||_2^2$$

is nearly equivalent to the analysis model (5.6)

$$\hat{\mathbf{x}}_a = A_2^+ (\min_{\mathbf{y}} \quad \lambda||A_2^+ \mathbf{y}||_1 + \frac{1}{2}||A_1 \mathbf{y} - \mathbf{b}||_2^2),$$

where A_2 has full column-rank, then naturally $A_2^+ = (A_2^T A_2)^{-1} A_2^T$ has full row-rank and $A_2^+ A_2 = I$.

Proof. The notation $\mathbf{z} = \mathbf{z}^{A_2} + \mathbf{z}^{A_2 \perp}$ is introduced to denote the decomposition

of a signal $\mathbf{z}$ to the part $\mathbf{z}^{A_2}$ in the column-span of A_2 and the part $\mathbf{z}^{A_2\perp}$ in the orthogonal subspace. We begin with the analysis model:

$$\begin{aligned}\hat{\mathbf{x}}_a &= A_2^+(\min_{\mathbf{y}^{A_2},\mathbf{y}^{A_2\perp}} \lambda||A_2^+(\mathbf{y}^{A_2}+\mathbf{y}^{A_2\perp})||_1+\frac{1}{2}||A_1(\mathbf{y}^{A_2}+\mathbf{y}^{A_2\perp})-(\mathbf{b}^{A_2}+\mathbf{b}^{A_2\perp})||_2^2)\\ &= A_2^+(\min_{\mathbf{y}^{A_2},\mathbf{y}^{A_2\perp}} \lambda||A_2^+\mathbf{y}^{A_2}+A_2^+\mathbf{y}^{A_2\perp}||_1+\frac{1}{2}||A_1\mathbf{y}^{A_2}-\mathbf{b}^{A_2}||_2^2+\frac{1}{2}||A_1\mathbf{y}^{A_2\perp}-\mathbf{b}^{A_2\perp})||_2^2)\end{aligned}$$

As $\mathbf{y}^{A_2\perp}$ is orthogonal to the columns of A_2, it is orthogonal to the rows of A_2^+. Thus we have $A_2^+\mathbf{y}^{A_2\perp}=0$, leading to:

$$\begin{aligned}\hat{\mathbf{x}}_a &= A_2^+(\min_{\mathbf{y}^{A_2},\mathbf{y}^{A_2\perp}} \lambda||A_2^+\mathbf{y}^{A_2}||_1+\frac{1}{2}||A_1\mathbf{y}^{A_2}-\mathbf{b}^{A_2}||_2^2+\frac{1}{2}||A_1\mathbf{y}^{A_2\perp}-\mathbf{b}^{A_2\perp})||_2^2)\\ &= A_2^+(\min_{\mathbf{y}^{A_2}} \lambda||A_2^+\mathbf{y}^{A_2}||_1+\frac{1}{2}||A_1\mathbf{y}^{A_2}-\mathbf{b}^{A_2}||_2^2+\min_{\mathbf{y}^{A_2\perp}}\frac{1}{2}||A_1\mathbf{y}^{A_2\perp}-\mathbf{b}^{A_2\perp})||_2^2)\end{aligned}$$

Signals $\mathbf{y}^{A_2}$ spanned by the columns of A_2 have a representation as $\mathbf{y}^{A_2}=A_2\mathbf{x}$. We can thus reformulate the former part of the above equation as an optimization on $\mathbf{x}$:

$$\begin{aligned}\hat{\mathbf{x}}_a &= A_2^+A_2(\min_{\mathbf{x}} \quad \lambda||A_2^+A_2\mathbf{x}||_1+\frac{1}{2}||A_1A_2\mathbf{x}-\mathbf{b}^{A_2}||_2^2)\\ &\quad +A_2^+(\min_{\mathbf{y}^{A_2\perp}} \quad \frac{1}{2}||A_1\mathbf{y}^{A_2\perp}-\mathbf{b}^{A_2\perp})||_2^2)\\ &= \min_{\mathbf{x}} \quad \lambda||\mathbf{x}||_1+\frac{1}{2}||A_1A_2\mathbf{x}-\mathbf{b}^{A_2}||_2^2\\ &\quad +A_2^+(\min_{\mathbf{y}^{A_2\perp}} \quad \frac{1}{2}||A_1\mathbf{y}^{A_2\perp}-\mathbf{b}^{A_2\perp})||_2^2)\\ &= \min_{\mathbf{x}} \quad \lambda||\mathbf{x}||_1+\frac{1}{2}||\mathbf{b}^{A_2}-A_1A_2\mathbf{x}||_2^2+\frac{1}{2}||\mathbf{b}^{A_2\perp}||_2^2\\ &\quad +A_2^+(\min_{\mathbf{y}^{A_2\perp}} \quad \frac{1}{2}||A_1\mathbf{y}^{A_2\perp}-\mathbf{b}^{A_2\perp})||_2^2)\\ &= \min_{\mathbf{x}} \quad \lambda||\mathbf{x}||_1+\frac{1}{2}||\mathbf{b}^{A_2}+\mathbf{b}^{A_2\perp}-A_1A_2\mathbf{x}||_2^2\\ &\quad +A_2^+(\min_{\mathbf{y}^{A_2\perp}} \quad \frac{1}{2}||A_1\mathbf{y}^{A_2\perp}-\mathbf{b}^{A_2\perp})||_2^2)\\ &= \min_{\mathbf{x}} \quad \lambda||\mathbf{x}||_1+\frac{1}{2}||A_1A_2\mathbf{x}-\mathbf{b}||_2^2\\ &\quad +A_2^+(\min_{\mathbf{y}^{A_2\perp}} \quad \frac{1}{2}||A_1\mathbf{y}^{A_2\perp}-\mathbf{b}^{A_2\perp})||_2^2)\end{aligned}$$

$$= \hat{\mathbf{x}}_s + A_2^+(\min_{\mathbf{y}^{A_2\perp}} \quad \frac{1}{2}||A_1\mathbf{y}^{A_2\perp} - \mathbf{b}^{A_2\perp})||_2^2)$$

Now we can conclude that the analysis solution and the synthesis solution are nearly equivalent and related by $\hat{\mathbf{x}}_a = \hat{\mathbf{x}}_s + A_2^+(\min_{\mathbf{y}^{A_2\perp}} \quad \frac{1}{2}||A_1\mathbf{y}^{A_2\perp} - \mathbf{b}^{A_2\perp})||_2^2)$, as claimed. To be more specific, the solving of the analysis model and the systhesis model is equivalent in the sense that one solution can be obtained from the other given fixed A_1, A_2 and $\mathbf{b}$. □

5.4 Optimization Algorithm – TVAL3

TVAL3 is short for 'TV minimization by Augmented Lagrangian and ALternating direction ALgorithms'. As the name suggests, it involves two important techniques. One is the augmented Lagrangian method, and the other is the alternating direction method (ADM). We start with the introduction of ADM.

5.4.1 General Framework of ADM

Suppose that $g(\mathbf{x}) : \mathbb{R}^p \to \mathbb{R}$ and $h(\mathbf{y}) : \mathbb{R}^q \to \mathbb{R}$ are convex functions, $A \in \mathbb{R}^{l\times p}, B \in \mathbb{R}^{l\times q}$ and $b \in \mathbb{R}^l$, the constrained optimization problem is formulated as:

$$\min_{\mathbf{x},\mathbf{y}} \quad g(\mathbf{x}) + h(\mathbf{y}) \quad \text{s.t.} \quad A\mathbf{x} + B\mathbf{y} = b. \tag{5.8}$$

Its augmented Lagrangian function is written as:

$$\mathcal{L}_\mathcal{A}(\mathbf{x}, \mathbf{y}, \boldsymbol{\lambda}) = g(\mathbf{x}) + h(\mathbf{y}) - \boldsymbol{\lambda}^T(A\mathbf{x} + B\mathbf{y} - b) + \frac{\alpha}{2}||A\mathbf{x} + B\mathbf{y} - b||_2,$$

where $\boldsymbol{\lambda} \in \mathbb{R}^l$ is the Lagrangian multiplier and $\alpha > 0$ is the penalty parameter.

In contrast to the classic augmented Lagrangian method, ADM takes advantage of the separable form of the objective function and minimizes $\mathcal{L}_\mathcal{A}(\mathbf{x}, \mathbf{y}, \boldsymbol{\lambda})$ with respect to $\mathbf{x}$ and $\mathbf{y}$ seperately via a Gauss-Seidel type iteration. ADM iterates as

follows: given $(\mathbf{y}^t, \boldsymbol{\lambda})$,

$$\begin{cases} \mathbf{x}^{t+1} \leftarrow & \arg\min_x \mathcal{L}_{\mathcal{A}}(\mathbf{x}, \mathbf{y}^t, \boldsymbol{\lambda}^t) \\ \mathbf{y}^{t+1} \leftarrow & \arg\min_y \mathcal{L}_{\mathcal{A}}(\mathbf{x}^{t+1}, \mathbf{y}, \boldsymbol{\lambda}^t) \\ \boldsymbol{\lambda}^{t+1} \leftarrow & \boldsymbol{\lambda}^t - \eta\alpha(A\mathbf{x}^{t+1} + B\mathbf{y}^{t+1} - b). \end{cases}$$

Here a step length η is introduced in the update of $\boldsymbol{\lambda}$, and it can guarantee the convergence of ADM when $\eta \in (0, (\sqrt{5}+1)/2)$ under certain technical assumptions [54, 55].

5.4.2 Specific Algorithm

The object is to solve:

$$\min_{\mathbf{x}} \quad \mathrm{TV}(\mathbf{x}) := \sum_i ||D_i\mathbf{x}||_2, \quad \text{s.t.} \quad A\mathbf{x} = \mathbf{b} \tag{5.9}$$

where $\mathbf{x} \in \mathbb{R}^N$ is the signal to be reconstructed, $D_i\mathbf{x}$ is the discrete gradient of $\mathbf{x}$ at pixel i, $A \in \mathbb{R}^{m\times N}(m < N)$ is the measurement matrix, and $\mathbf{b}$ denotes the measurement. The regularization term is addressed as the isotropic TV norm and the developed algorithm can be easily extended to solve the problem with the anisotropic TV norm, which is $\mathrm{TV_a}(\mathbf{x}) := \sum_i ||D_i\mathbf{x}||_1$.

(5.9) is equivalent to:

$$\min_{\mathbf{w}_i, \mathbf{x}} \quad \sum_i ||\mathbf{w}_i||_2 \quad \text{s.t.} \quad A\mathbf{x} = \mathbf{b}, D_i\mathbf{x} = \mathbf{w}_i, \forall i \tag{5.10}$$

the corresponding augmented Lagrangian function of which is:

$$\begin{aligned} \mathcal{L}_{\mathcal{A}}(\mathbf{w}_i, \mathbf{x}, \mathbf{v}_i, \boldsymbol{\lambda}) &= \sum_i (||\mathbf{w}_i||_2 - \mathbf{v}_i^T(D_i\mathbf{x} - \mathbf{w}_i) + \frac{\alpha_i}{2}||D_i\mathbf{x} - \mathbf{w}_i||_2^2) \\ &\quad -\boldsymbol{\lambda}(A\mathbf{x} - \mathbf{b}) + \frac{\tau}{2}||A\mathbf{x} - \mathbf{b}||_2^2. \end{aligned} \tag{5.11}$$

$\mathbf{v}_i, \boldsymbol{\lambda}$ here are the Lagrangian multipliers.

Given $\mathbf{x}^t, \mathbf{v}_i^t$ and $\boldsymbol{\lambda}^t$, the minimizer $\mathbf{w}_i^{t+i}$ is updated as follows:

$$\mathbf{w}_i^{t+i} = \mathrm{Shrink2}(D_i\mathbf{x}^t - \mathbf{v}_i/\alpha_i, \frac{1}{\alpha_i})$$

$$:= \max\{||D_i\mathbf{x}^t - \mathbf{v}_i/\alpha_i||_2 - \frac{1}{\alpha_i}, 0\}\frac{D_i\mathbf{x}^t - \mathbf{v}_i/\alpha_i}{||D_i\mathbf{x}^t - \mathbf{v}_i/\alpha_i||_2}.$$

In the case of anisotropic TV norm, the update of $\mathbf{w}_i^{t+1}$ is given by:

$$\begin{aligned}\mathbf{w}_i^{t+i} &= \text{Shrink1}(D_i\mathbf{x}^t - \mathbf{v}_i/\alpha_i, \frac{1}{\alpha_i}) \\ &:= \max\{|D_i\mathbf{x}^t - \mathbf{v}_i/\alpha_i| - \frac{1}{\alpha_i}, 0\} \circ \text{sgn}(D_i\mathbf{x}^t - \mathbf{v}_i/\alpha_i).\end{aligned}$$

where $\circ$ denotes component-wise multiplication, and $\text{sgn}(z) := \begin{cases} -1 & \text{if } z < 0 \\ 0 & \text{if } z = 0 \\ 1 & \text{if } z > 0 \end{cases}$.
Interested readers can refer to [91] pp 21-25 for details and proof.

With fixed $\mathbf{w}_i^{t+1}, \mathbf{v}_i^t$ and $\boldsymbol{\lambda}^t$, (5.11) is quadratic with respect to $\mathbf{x}$ and its gradient can be easily derived as:

$$\mathbf{d}^t(\mathbf{x}) = \sum_i(\alpha_i D_i^T(D_i\mathbf{x} - \mathbf{w}_i^{t+1}) - D_i^T\mathbf{v}_i) + \tau A^T(A\mathbf{x} - \mathbf{b}) - A^T\boldsymbol{\lambda}.$$

It is ideal to force $\mathbf{d}^t(\mathbf{x}) = 0$ for the exact minimizer $\mathbf{x}^{t+1}$, however, the numerical implement is too costly to realise. A one-step steepest descent method is proposed here instead:

$$\mathbf{x}^{t+1} = \mathbf{x}^t - \epsilon^t\mathbf{d}^t(\mathbf{x}^t),$$

where ϵ^t is the step length chosen in the aggressive manner proposed by Barzilai and Borwein [8], known as the BB step. It is validated by the nonmonotone Armijo condition [166].

Finally, the multipliers are updated by:

$$\begin{aligned}\mathbf{v}_i^{t+1} &= v_i^t - \eta\alpha_i(D_i\mathbf{x}^{t+1} - \mathbf{w}_i^{t+1}), \forall i \\ \boldsymbol{\lambda}^{t+1} &= \boldsymbol{\lambda}^t - \eta\tau(A\mathbf{x}^{t+1} - \mathbf{b})\end{aligned}$$

In conclusion, (5.9) is solved via ADM with iterations as below:

$$\begin{cases} \mathbf{w}_i^{t+i} & = \text{Shrink2}(D_i\mathbf{x}^t - \mathbf{v}_i/\alpha_i, \frac{1}{\alpha_i}), \forall i \\ \mathbf{x}^{t+1} & = \mathbf{x}^t - \epsilon^t \mathbf{d}^t(\mathbf{x}^t) \\ \mathbf{v}_i^{t+1} & = v_i^t - \eta\alpha_i(D_i\mathbf{x}^{t+1} - \mathbf{w}_i^{t+1}), \forall i \\ \boldsymbol{\lambda}^{t+1} & = \boldsymbol{\lambda}^t - \eta\tau(A\mathbf{x}^{t+1} - \mathbf{b}) \end{cases}$$

The convergence analysis is provided in [92] pp 23-31.

5.5 Simulation

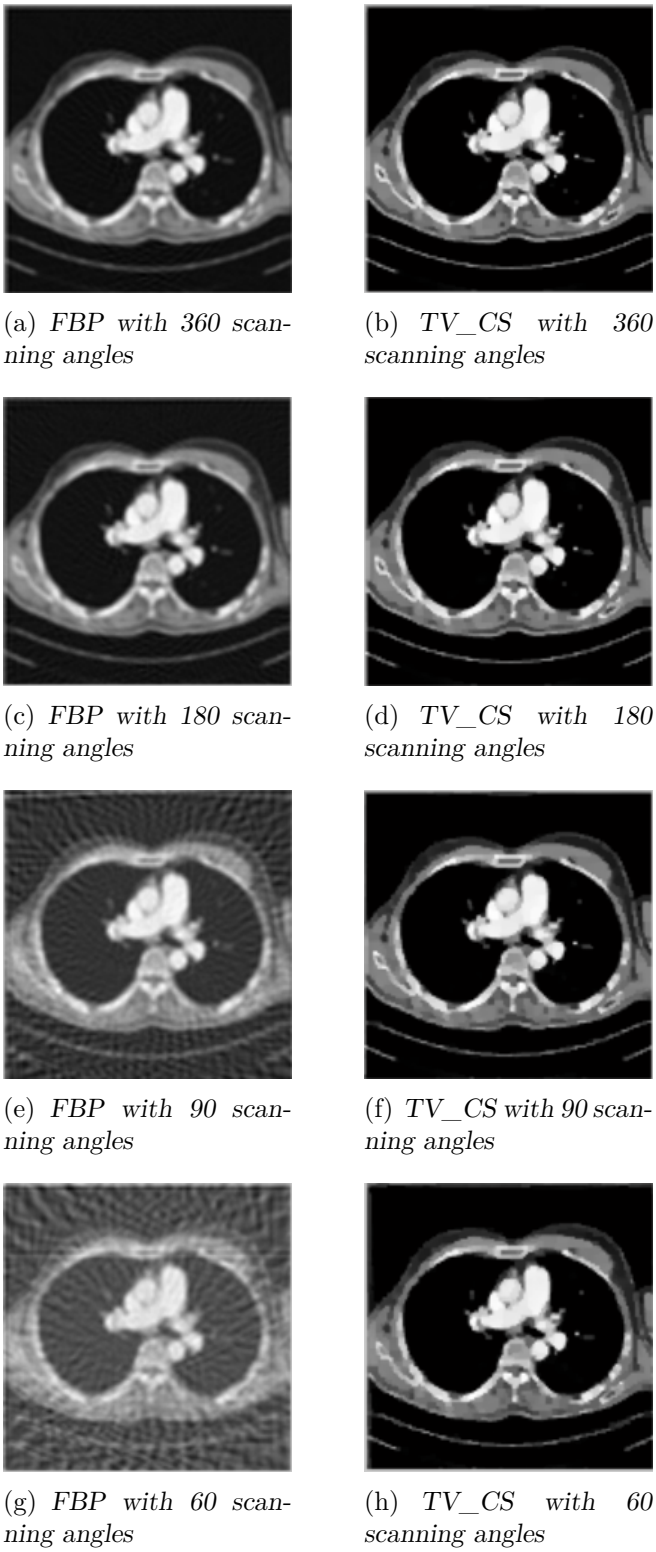
5.5.1 Comparison between FBP and TV-based CS

The experiment was conducted over twenty real CT images, and two were selected for illustration here. The image size was set to be 128×128 and the distance from the source to the image centre equalled the image size in each dimension, i.e. $R = n$. The spacing of the fan-beam sensors was $0.7°$, and the number of sensors was 133, determined by calculating how many beams were required to cover the entire image for any rotation angle. A complete projection set should be taken at $1°$ increments over a full $360°$ range, and the low-radiation projections were obtained with the golden-angle scanning, corresponding to $50\%, 25\%, 17\%$ of the full projections, respectively. For the TV-based CS method, the results are shown in Figure 5.1 and Figure 5.2.

As the figures show, with the decrease of projections, the reconstruction quality of FBP also decreases, while CS manages to maintain good reconstruction until the 17% of the full projections. In order to further validate our analysis, we assessed the performance of the reconstructions from different sampling patterns. The normalized mean square error (NMSE) was adopted to indicate the closeness between the original image f_0 and the recovered image f:

$$\text{NMSE} = \mathbb{E}\left\{ \frac{||f - f_0||_2}{||f_0||_2} \right\}$$

The numerical results of CT1 and CT2 are shown in Table 5.1. They are consistent with the visual results.



(a) *FBP with 360 scanning angles*

(b) *TV_CS with 360 scanning angles*

(c) *FBP with 180 scanning angles*

(d) *TV_CS with 180 scanning angles*

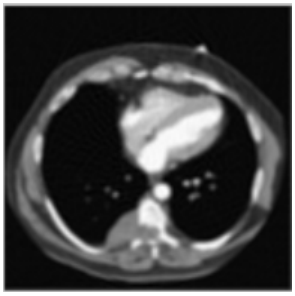
(e) *FBP with 90 scanning angles*

(f) *TV_CS with 90 scanning angles*

(g) *FBP with 60 scanning angles*

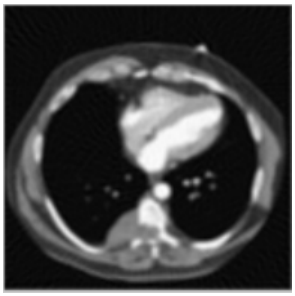
(h) *TV_CS with 60 scanning angles*

Figure 5.1: Comparison of reconstruction results on CT1 from FBP and TV_CS



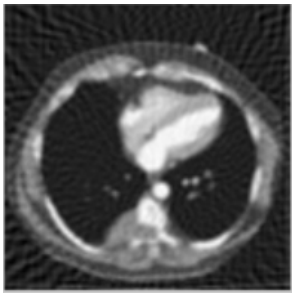
(a) *FBP with 360 scanning angles*

(b) *TV_CS with 360 scanning angles*



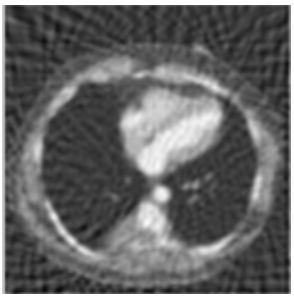
(c) *FBP with 180 scanning angles*

(d) *TV_CS with 180 scanning angles*



(e) *FBP with 90 scanning angles*

(f) *TV_CS with 90 scanning angles*



(g) *FBP with 60 scanning angles*

(h) *TV_CS with 60 scanning angles*

Figure 5.2: Comparison of reconstruction results on CT2 from FBP and TV_CS

Table 5.1: Comparison of reconstruction results in terms of NMSE

	CT1				CT2			
No. of scanning angles	360	180	90	60	360	180	90	60
FBP	0.18	0.19	0.26	0.32	0.13	0.14	0.19	0.25
TV_CS	0.09	0.11	0.12	0.16	0.08	0.09	0.10	0.11

5.5.2 CS-Based Reconstruction with Different Sampling Schemes

The experimental setting was as follows: The image size was set to be 128×128. The distance from the source to the centre of the domain was $R = 2n$. The span of the rays was defined such that from $(2n, 0)$ the first ray hit the point $(n/2, n/2)$ and the last ray hit $(n/2, -n/2)$. We took the reference point of full sampling with $2n$ samples in both the view and bin direction, i.e. $N_{\text{views}} = N_{\text{bins}} = 2n$. Two groups of simulations were designed to test the detector and scan angle sampling, respectively. For the first group, the view number was 256. The detector sampling ratio varied from $10\% - 80\%$, with regular and random distribution. For the second group, the detector number was 256. The angle sampling ratio varied from $10\% - 80\%$, with golden-angle and random-angle patterns. In the random sampling case, the values were averaged over 10 groups of experiments. The results are recorded in Table 5.2 and Table 5.3, respectively, where μ is the sidelobe-to-peak ratio (SPR) developed in [94].

Table 5.2: Group1 the detector setting

	regular-detector		random-detector	
sampling ratio	$\Upsilon(10^2)$	$\mu(10^{-1})$	$\Upsilon(10^2)$	$\mu(10^{-1})$
10%	3.75	8.61	3.11	6.69
20%	2.97	9.47	2.73	6.23
30%	2.68	7.65	2.57	5.84
40%	2.55	6.78	2.50	5.63
50%	2.45	6.78	2.44	5.53
60%	2.44	6.15	2.42	5.50
70%	2.41	5.70	2.39	5.37
80%	2.38	5.40	2.37	5.29

According to Table 5.2 and Table 5.3, golden-angle and random-detector obtain

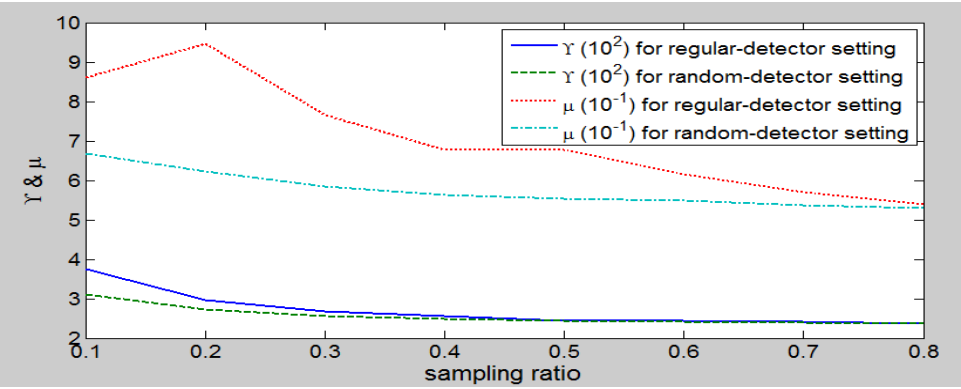
Table 5.3: Group2 the angle setting

	golden-angle		random-angle	
sampling ratio	$\Upsilon(10^2)$	$\mu(10^{-1})$	$\Upsilon(10^2)$	$\mu(10^{-1})$
10%	3.05	5.67	3.27	6.08
20%	2.61	5.35	2.81	5.62
30%	2.48	5.37	2.61	5.42
40%	2.42	5.34	2.53	5.45
50%	2.39	5.31	2.47	5.38
60%	2.38	5.29	2.45	5.35
70%	2.37	5.18	2.40	5.27
80%	2.36	5.24	2.39	5.25

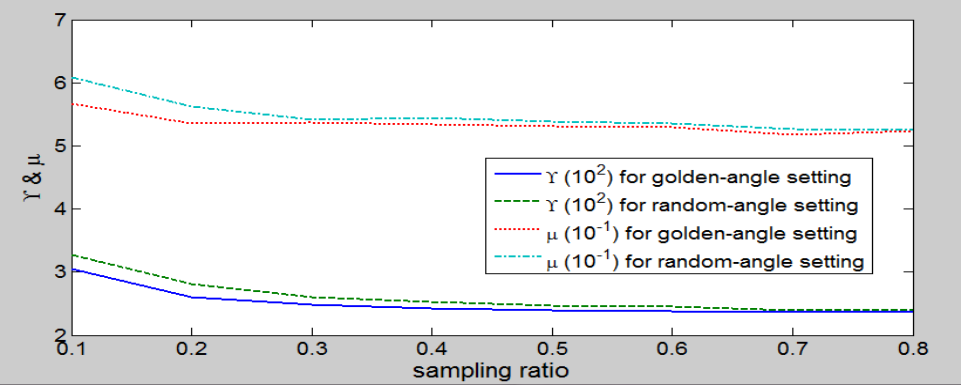
lower Υ and μ than random-angle and regular-detector in under-sampled cases, respectively, indicating higher incoherence of the system matrix and better reconstruction through CS. It can also be seen from Table 5.2 and Table 5.3 that the index developed in this thesis is more sensitive than that in [94] because Υ monotonically decreases with the increase of the sampling ratio, while μ shows some exceptions. The results are also shown graphically in Figure 5.3.

Two classes of images were involved as the object, artificial phantom images and real CT images. Twenty images were taken from each of class, and three were selected to show the reconstruction results, as in Figure 5.4 and Table 5.4. The reconstructions with different sampling patterns are displayed in Figure 5.5, Figure 5.6 and Figure 5.7.

Table 5.4 shows that the NMSE is lower with the random-detector and regular-angle sampling schemes than the other two, especially in the heavily under-sampled case. Putting Table 5.2 and Table 5.3 together with Table 5.4, we can see that the reduction of Υ for 10^2 or the reduction of μ for 10^{-1} contributes to the reduction of the NMSE for at least 10%. We can also see from Figure 5.5, Figure 5.6 and Figure 5.7 that the reconstructed images from the regular-detector and random-angle schemes show more noise and blurred details. Hence, we can conclude that simulation results are consistent with the theory prediction, that the random-detector and regular-angle schemes provide more satisfactory reconstructions than the regular-detector and random-angle schemes.



(a) *the detector setting*



(b) *the angle setting*

Figure 5.3: PSF variations with regard to the sampling ratio

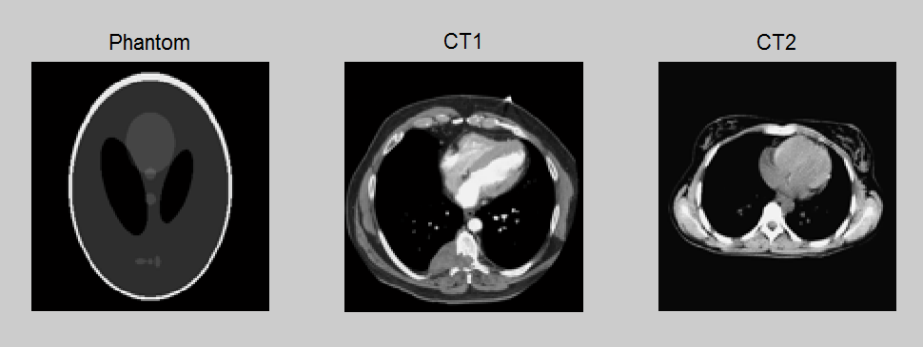


Figure 5.4: Reconstruction objects

5.6 Conclusion

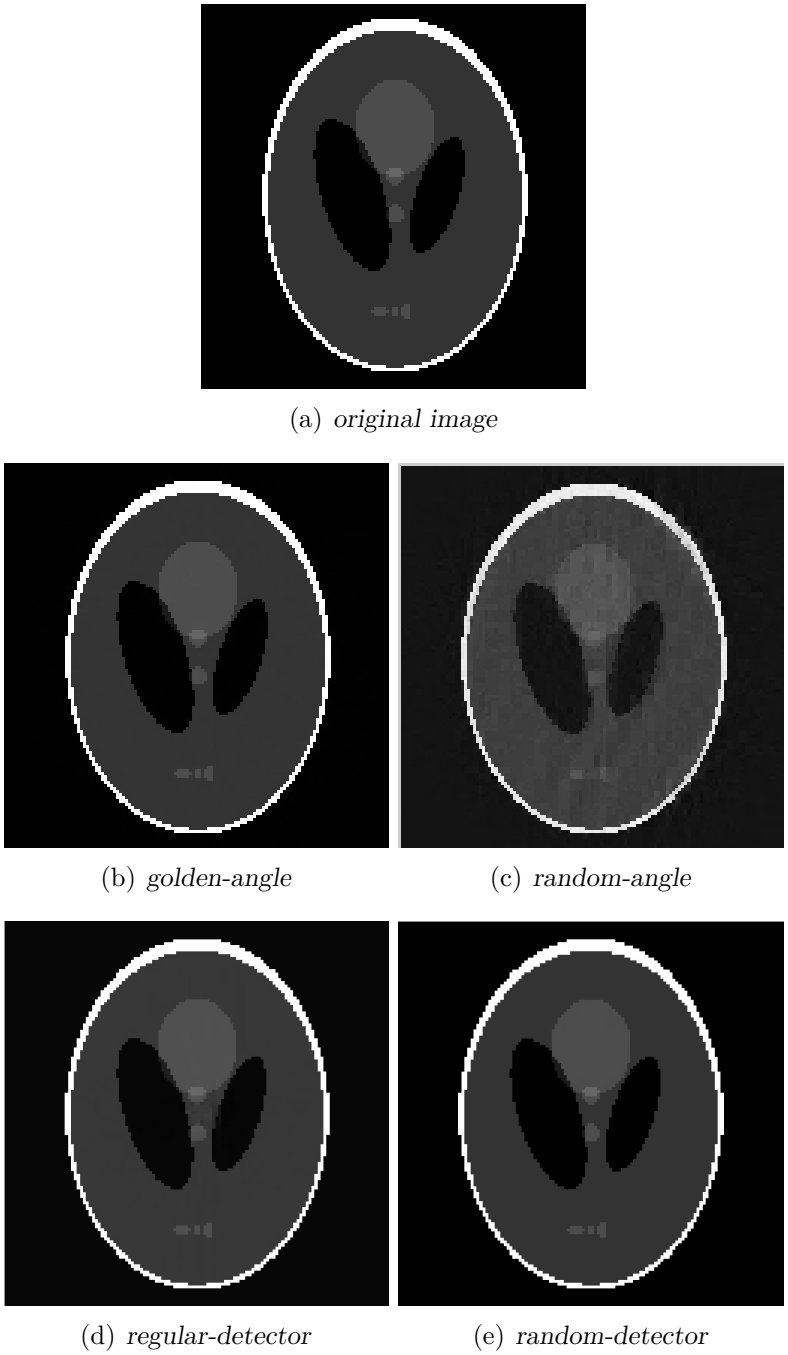
In this chapter, the property of the system matrix for fan-beam CT has been investigated with the purpose of both reducing the radiation dose and preserving high-quality recovery. With the under-sampled projections, CS was employed for the recovery using the optimization algorithm of TVAL3, and has shown

Table 5.4: NMSE with different sampling patterns in fan-beam CT

Shepp-logan	10%	20%	30%	40%	50%	60%	70%	80%
regular-angle	10.18%	0.20%	0.13%	0.12%	0.11%	0.10%	0.11%	0.09%
random-angle	22.01%	3.98%	0.15%	0.13%	0.12%	0.12%	0.12%	0.11%
regular-detector	40.55%	8.24%	1.31%	0.43%	0.17%	0.12%	0.13%	0.10%
random-detector	11.23%	0.16%	0.13%	0.12%	0.12%	0.11%	0.11%	0.11%
CT1	10%	20%	30%	40%	50%	60%	70%	80%
regular-angle	22.22%	6.91%	0.90%	0.26%	0.18%	0.15%	0.13%	0.13%
random-angle	24.60%	9.33%	3.23%	0.88%	0.32%	0.19%	0.15%	0.14%
regular-detector	39.58%	16.63%	6.44%	1.92%	0.56%	0.22%	0.15%	0.14%
random-detector	17.67%	5.97%	0.57%	0.21%	0.17%	0.15%	0.14%	0.13%
CT2	10%	20%	30%	40%	50%	60%	70%	80%
regular-angle	22.19%	6.13%	0.61%	0.16%	0.15%	0.12%	0.12%	0.11%
random-angle	25.55%	10.15%	3.37%	0.33%	0.27%	0.15%	0.13%	0.13%
regular-detector	41.14%	17.65%	6.23%	1.46%	0.50%	0.26%	0.14%	0.12%
random-detector	18.17%	5.04%	0.55%	0.17%	0.15%	0.13%	0.12%	0.11%

advantages over FBP in the simulations. Four under-sampling patterns were considered: golden-angle, random-angle, regular-detector and random-detector. Two tools were used for the analysis of the impacts of different sampling schemes on reconstructions: PSF and FST. Based on PSF, an evaluation index was proposed: the Frobenius norm Υ of the difference matrix between PSF and the identity matrix. The lower Υ, the better the reconstruction. In contrast, high off-diagonal interference in PSF implied low-quality recovery. Based on FST, the system matrix was decomposed to Fourier matrix and our aim was converted to the study of sampling on Fourier matrix. The golden-angle system obtained lower Υ than random-angle and the random-detector setting obtained lower Υ than regular-detector. Experiments were conducted on phantom image and real CT images, of which the results were measured by normalized mean square error (NMSE). NMSE results were consistent with the indicator results, where golden-angle scanning and random-detector setting obtained lower NMSE than the other two, respectively.

Using both the theoretical and experimental analysis, the purpose of reducing radiation dose without compromising the reconstruction quality has been achieved. The conclusion is that golden-angle and random-detector scanning is the most



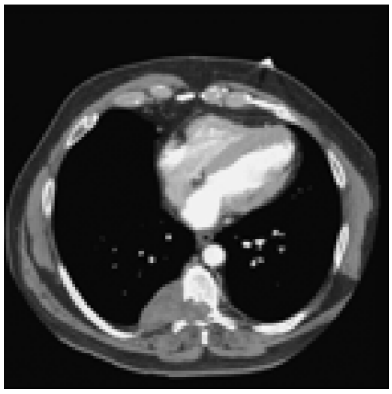
(a) *original image*

(b) *golden-angle* (c) *random-angle*

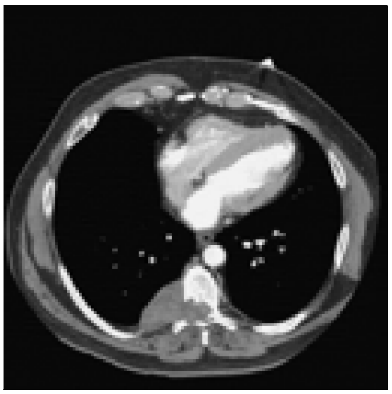
(d) *regular-detector* (e) *random-detector*

Figure 5.5: Phantom reconstruction from 10% angle and 30% detector

favourable for post-CS-based reconstruction.



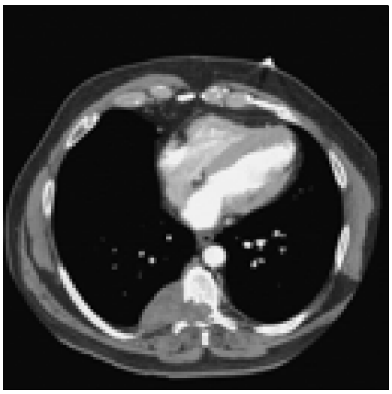
(a) *original image*



(b) *golden-angle*

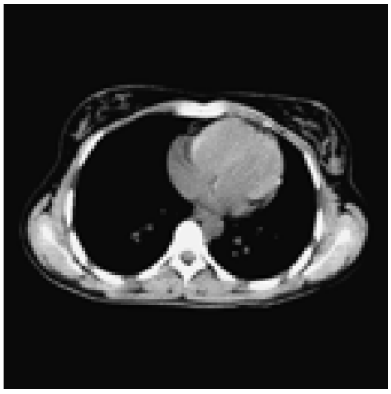
(c) *random-angle*

(d) *regular-detector*

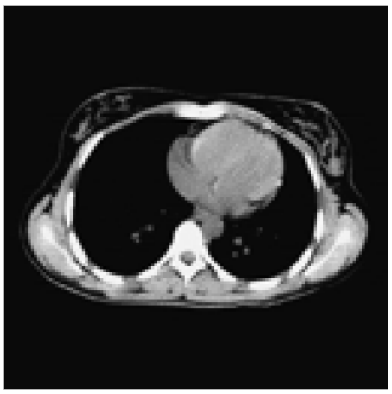


(e) *random-detector*

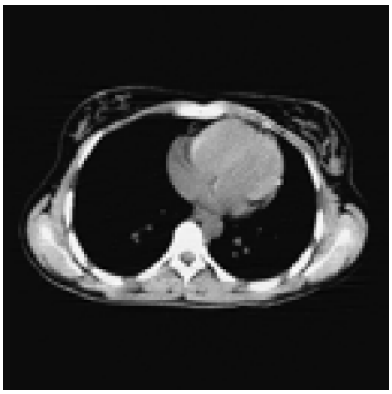
Figure 5.6: CT1 reconstruction from 30% angle and 30% detector



(a) *original image*

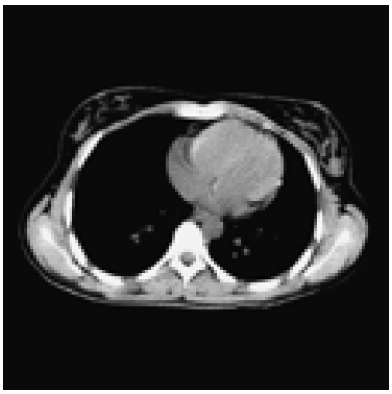


(b) *golden-angle*



(c) *random-angle*

(d) *regular-detector*



(e) *random-detector*

Figure 5.7: CT2 reconstruction from 30% angle and 30% detector

Chapter 6

Compressed Sensing-Based CT Reconstruction from Mojette Projections

6.1 Introduction

Classical tomography is concerned with the recovery of the scanned object from a set of projections, which is the Radon transform (RT) [146] of the object. As Radon transform is continuous, both for the object under scan and the projections themselves, it has to be sampled to adjust to practical applications, resulting in ill-posedness. The conventional solution is to acquire the projections over a large number of view angles and use filtering in the frequency domain, which on one hand increases radiation dose and on the other still produces artifacts in the reconstruction result. Discrete Radon transform gives a different view of the tomographic problem. It regularizes the ill-posedness of inverse Radon transform and allows for an exact reconstruction in the discrete domain with a finite number of projections. Many discrete Radon methods, for example, arbitrary curve block circulant discrete RT [12, 84], d-lines discrete RT [17, 58] and Fast Slant Stack [6], tend to suffer from either experimental limitations or high computational complexity [28]. This chapter will focus on the Mojette transform, first proposed by Guédon et al. [59], which is a both experimentally and computationally viable discretization of Radon transform. Despite its advantages, Mojette

transform has two main shortcomings. One is that it is not compatible with the physical acquisition of CT, and this problem is addressed in [127], in which the authors design a linear system to calculate the Mojette projections from a Radon acquisition. The other is that the angle and bin set are different from practical CT machine, i.e., classical tomography acquires are equally distributed over 2π with a fixed number of bins onto each projection, while Mojette transform is defined over Farey angles with varying orientation and number of bins onto each projection. Much research [27,48,85] has been done, committed to applying Mojette transform to classical tomographic data, which implies the advantage of Mojette transform on one hand and improves the potential of Mojette transform in practical applications on the other hand. Fayad et al. [48] claim that any set of real, acquired tomographic data can be rebinned into a compatible Mojette projection space, without any loss of reconstruction power.

After obtaining the projections, the next step is the reconstruction problem. Katz [83] presents some reconstructibility theorems for discrete images and projections. A good deal of work has also been done to recover objects based on Mojette projections, committed to the improvement of accuracy. Since we also attempt to reduce the radiation dose, a newly developed technique called compressed sensing (CS) is used. CS attempts to reconstruct signals from significantly fewer samples than were traditionally thought necessary (cf. the Nyquist sampling theorem) [95]. CT imagery meets the two key requirements for successful application of CS: it is compressible in some transform domain and the scanner receives encoded samples. Applying CS to CT reconstruction, high-accuracy reconstruction with low-dose radiation is expected to be obtained [61].

The main contribution of the chapter is that a novel reconstruction frame, i.e. the combination of Mojette transform and CS, is proposed in the interests of accuracy and reduced radiation. A detailed description and explanation of how to modify Mojette projections for the application of CS is given. In addition, the reason for using Mojette transform instead of Radon transform is analyzed. We also show a simple way to calculate the point spread function (PSF) of the proposed method by proving the PSF should be a Toeplitz matrix. The rest of the chapter is organized as follows. In Section 6.2, the Mojette transform and CS is briefly introduced, and the relation between Mojette transform and the 2-D Fourier coefficients of the scanning object is formulated. Next, the detailed reconstruction frame is described and thoroughly compared with the direct application of CS.

Finally, the specified optimization algorithm of RecPF is described. Section 6.3 gives the experimental results to validate the accuracy, low-dosage radiation and noise tolerance of the proposed frame, and also presents some discussion. Conclusions are drawn in Section 6.4.

6.2 Reconstruction Frame Based on Mojette Transform and Compressed Sensing

6.2.1 Mojette Transform

For simplicity, we assume the scanning object to be $f(x, y)$ with the size $n \times n$ and n is even. The CT scanner maps the 2-D object into a set of 1-D projection lines, which forms the sinogram. Each line is the Radon transform of f for a given angle θ and a module r [146], defined by:

$$s(\theta, r) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\delta(r - x\cos\theta - y\sin\theta)dxdy \tag{6.1}$$

where θ and r are respectively the angular and radial coordinates of the projection line (θ, r), and $\delta(\cdot)$ is the Dirac delta function.

The Dirac-Mojette transform is an exact discretization of the Radon transform. It is defined over angles $\theta = \tan^{-1}(\frac{q}{p})$ where p and q are relatively prime integers. As the Farey series of order K, denoted by F_K, is the set of all fractions in lowest terms between 0 and ∞ whose denominators do not exceed K, e.g. $F_4 = [\frac{0}{1}, \frac{1}{4}, \frac{1}{3}, \frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \frac{1}{1}, \frac{4}{3}, \frac{3}{2}, \frac{2}{1}, \frac{3}{1}, \frac{4}{1}, \frac{1}{0}]$, we can use it to give a set of discrete angles between $[0, \frac{\pi}{2}]$ and obtain the angles over $[0, \pi]$ by symmetry. The Dirac-Mojette transform is defined as:

$$M_{p,q}(r) = \sum_x \sum_y f(x, y)\delta(r - px - qy) \tag{6.2}$$

The geometry of the Dirac-Mojette projector is shown in Figure 6.1, where the pixel is summed to its corresponding bin if and only if the X-ray passes through the centre of the pixel [28].

Fourier slice theorem claims that the 1-D Fourier transform of the projections is

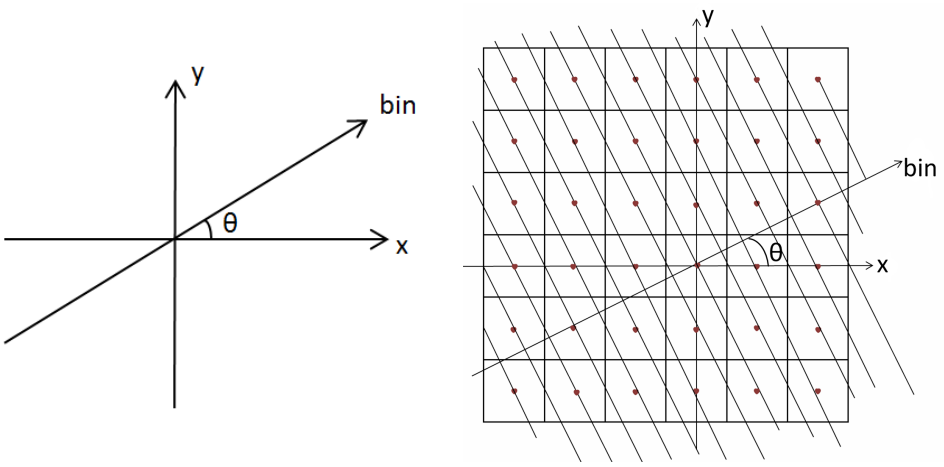


Figure 6.1: Dirac-Mojette Projection of a 6×6 image

equal to the 2-D Fourier transform of the image evaluated on the line that the projection was taken on. In this chapter, it is formulated that the 2-D Fourier values of the scanning object can be obtained exactly through Mojette transform. Suppose (u, v) is the coordinate in the frequency domain, then the 2-D discrete Fourier transform of $f(x, y)$ is:

$$\mathcal{F}(u, v) = \sum_x \sum_y f(x, y) e^{-j\frac{2\pi}{n}(ux+vy)} \tag{6.3}$$

The 1-D discrete Fourier transform of $M_{p,q}(r)$ is:

$$\mathcal{FM}_{p,q}(\rho) = \sum_\rho M_{p,q}(r) e^{-j\frac{2\pi}{L}\rho r} \tag{6.4}$$

where L is the projection size, and $L = (n-1)(|p| + |q|) + 1$. Combining (6.2) and (6.4), we obtain:

$$\mathcal{FM}_{p,q}(\rho) = \sum_x \sum_y f(x, y) e^{-j\frac{2\pi}{L}\rho(px+qy)} \tag{6.5}$$

Comparing (6.3) and (6.5), it is hard to map $\mathcal{FM}_{p,q}(r)$ to $\mathcal{F}(u, v)$ directly. Considering the periodicity of the Fourier transform, we can merge the items gained by (6.2) with the same $mod(px + qy, n)$, then projection size can be reduced to n. So

$$M'_{p,q}(r) = \sum_x \sum_y f(x, y)|_{r=mod(px+qy,N)} \tag{6.6}$$

Its 1-D Fourier transform is:

$$\begin{aligned}\mathcal{FM}'_{p,q}(\rho) &= \sum_x \sum_y f(x,y) e^{-j\frac{2\pi}{n}\rho[mod(px+qy,n)]} \\ &= \sum_x \sum_y f(x,y) e^{-j\frac{2\pi}{n}[mod(\rho px+\rho qy,n)]}\end{aligned} \tag{6.7}$$

The mapping relationship is revealed by the analysis of (6.3) and (6.7):

$$\forall x, y, mod(ux + vy - rpx - rqy, n) = 0 \tag{6.8}$$

Hence, the 1-D Fourier transform of the modified Mojette projections taken over angles $\theta = \tan^{-1}(\frac{q}{p})$ can be mapped to the 2-D Fourier plane of the scanning object with the mapping relationship:

$$mod(u - rp, n) = 0 \qquad \& \qquad mod(v - rq, n) = 0 \tag{6.9}$$

As a result, we can obtain the partial Fourier coefficients where $mod(qu - pv, n) = 0$. The Fourier coefficients gained through the mapping of the Mojette projections in Figure 6.1 ($n = 6, tan\theta = \frac{1}{2}$) are marked with stars in Figure 6.2.

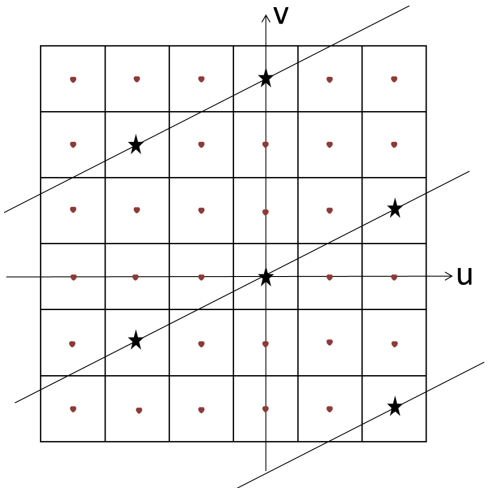


Figure 6.2: Corresponding Fourier values of the Mojette projections in Figure 6.1

6.2.2 Formulation of CS-Based Reconstruction Model

Compressed sensing is a recently developed theory of signal recovery from highly incomplete information, and is employed here to reconstruct the image from the partial Fourier coefficients. The central idea of CS is that a sparse or compressible signal $\mathbf{x} \in \mathbb{C}^N$ can be recovered from a small number of linear measurements $\mathbf{b} = A\mathbf{x} \in \mathbb{C}^m, m \ll N$.

The reconstruction can be achieved by solving the well-known basis pursuit problem [23]:

$$\min \quad ||\mathbf{x}||_1 \quad \text{s.t.} \quad A\mathbf{x} = \mathbf{b} \tag{6.10}$$

With noisy and incomplete samples, an appropriate relaxation is given by:

$$\min \quad ||\mathbf{x}||_1 \quad \text{s.t.} \quad ||A\mathbf{x} - \mathbf{b}||_2 \leq \epsilon \tag{6.11}$$

where $\epsilon > 0$ is related to the noise. To gain more effective reconstruction results, a simple and fast algorithm called RecPF (Reconstruction from Partial Fourier data [160] is adopted. It uses an alternating minimization scheme to solve the following model:

$$\min_{\mathbf{x}} \quad \text{TV}(\mathbf{x}) + \tau_1||\varphi\mathbf{x}||_1 + \frac{\tau_2}{2}||F_p\mathbf{x} - \hat{\mathbf{x}}_p||^2 \tag{6.12}$$

where $\text{TV}(\mathbf{x})$ is the total variation regularization term, φ is a sparsifying basis (e.g., wavelet basis), F_p is a partial Fourier matrix and $\hat{\mathbf{x}}_p$ denotes the partial Fourier coefficients. As the main computation in solving the model only involves shrinkage and fast Fourier transforms, the reconstruction process is quite fast.

It is noticed that when the system matrix is a partial Fourier matrix, i.e. $A = F_p$, its point spread function (PSF) possesses certain properties and can be calculated efficiently. For simplicity, we investigate the 1-D case. (6.12) implies that:

$$\hat{\mathbf{x}}_p = F_p\mathbf{x},$$

and it leads to

$$\mathbf{x}_p = F_p^* F_p\mathbf{x},$$

where $x_p = F_p^*\hat{\mathbf{x}}_p$. This shows that PSF here is the relationship between $\mathbf{x}_p$ and $\mathbf{x}$.

Suppose

$$\hat{\mathbf{x}}'_p = \hat{\mathcal{P}}\hat{\mathbf{x}}, \tag{6.13}$$

where $\hat{\mathcal{P}}$ is a $N \times N$ diagonal matrix with m random 1 and $N - m$ 0 along the diagonal line, we have

$$\mathbf{x}_p = F^* \hat{\mathcal{P}} F \mathbf{x},$$

where F stands for the full Fourier matrix, and its conjugate F^* the inverse Fourier matrix.

Now we can conclude that

$$\text{PSF} = F_p^* F_p = F^* \hat{\mathcal{P}} F \mathbf{x}.$$

This makes the calculation of PSF quite efficient as F and F^* can be implemented via operations rather than specific matrices.

On the other hand, due to the complex convolution property of discrete Fourier transform, it is deduced from (6.13) that

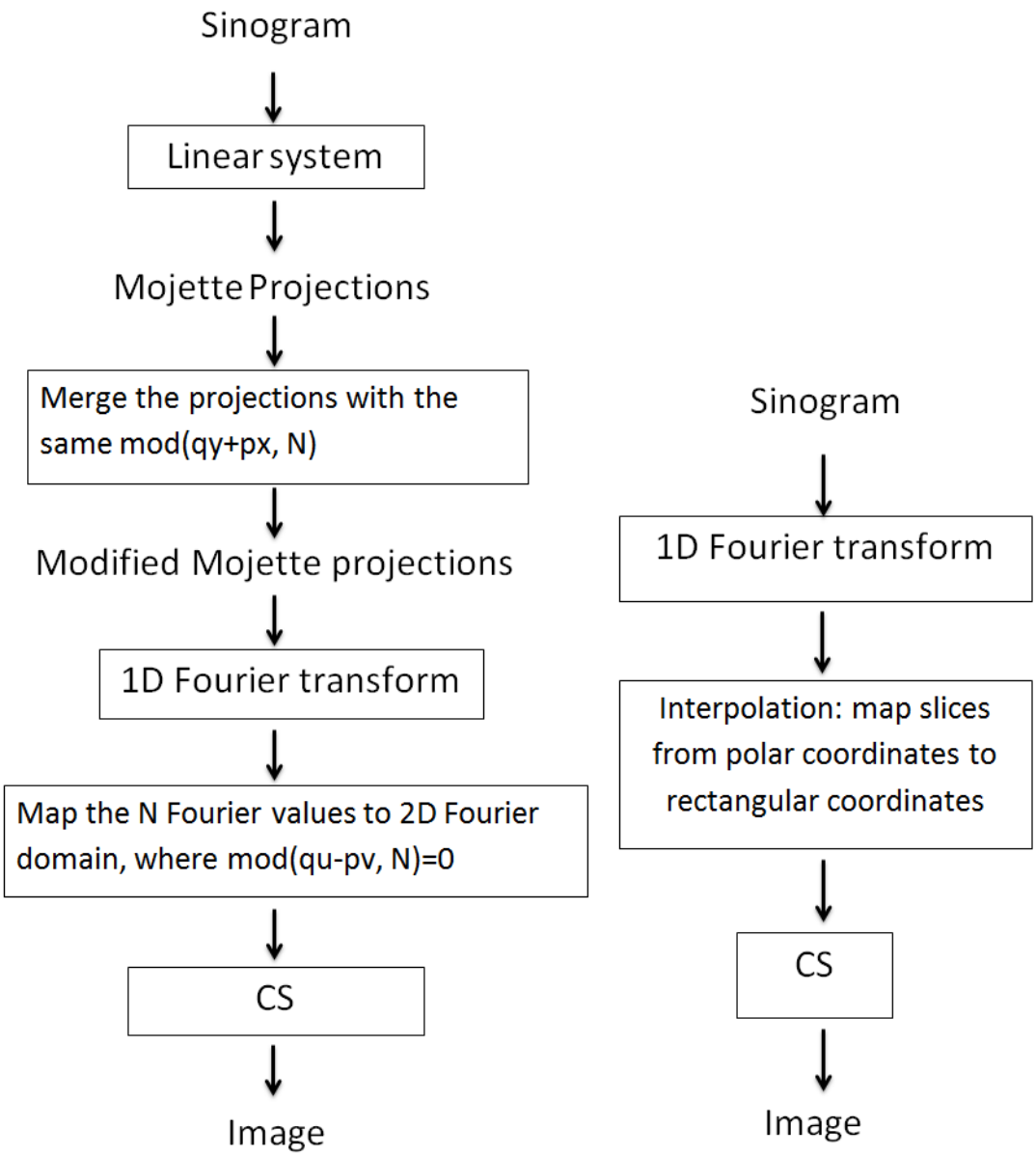
$$\mathbf{x}_p = \mathcal{P} \underset{N}{*} \mathbf{x},$$

where $\mathcal{P}$ is the inverse Fourier transform of $\hat{\mathcal{P}}$, and $\underset{N}{*}$ denotes circular convolution. Hence, the PSF must be a Toeplitz matrix. The proof can be easily applied to the 2-D case.

6.2.3 Proposed Frame

The proposed frame contains two stages. In the first stage, the sinogram is converted to Mojette projections using the linear system designed in [127], then their Fourier values are mapped to the 2-D Fourier domain of the object. In the second stage, CS is applied to the partial Fourier coefficients for the recovery of the object. In this section, details are given regarding the advantages of the proposed method (MCS) over the direct CS application in CT reconstruction, the flowcharts for which are shown in Figure 6.3, respectively.

As we can see from the flowcharts, using Mojette projections instead of Radon projections can avoid different gridding problems. The gridding problem is ad-



(a) *MCS* (b) *direct CS*

Figure 6.3: Flowcharts of MCS and direct CS

dressed in Chapter 4, which stated that the error in Fourier domain reduces with the increase of the zero-padding size. It involves large computation load but not complete noise removal, while the mapping is theoretically exact in the Fourier domain from Mojette projections.

In addition, Radon transform has the problem of irregular sampling, as shown in Fig. 6.4. In the scheme at left, the grey pixel is omitted in the sampling while in

the scheme at right , the middle grey pixel is considered twice while others are considered only once (the dark grey and light grey pixels). The irregular sampling also causes artifacts in the Fourier coefficients.

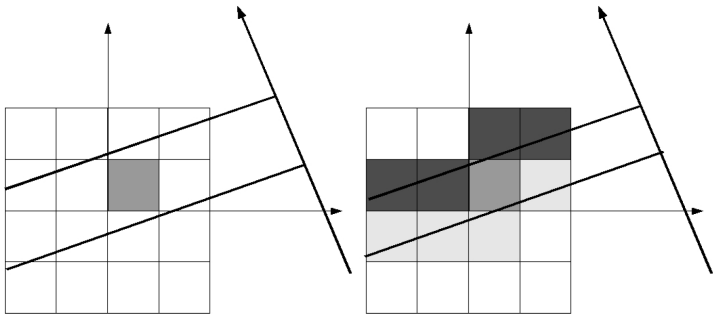


Figure 6.4: Irregular sampling of Radon transform

According to the above analysis, we know that dealing with Radon projections directly results in very noisy Fourier coefficients. In contrast, Mojette transform overcomes the irregular sampling of the Radon transform, solves the gridding problem and thus obtains the exact Fourier values. Combining it with CS, it is expected to obtain high-accuracy reconstruction results from limited-angle projections (i.e. low-dosage radiation).

6.2.4 Optimization Algorithm – RecPF

RecPF is a fast alternating direction method for TVL1-L2 signal reconstruction from partial Fourier data proposed by Yang et al. [160]. The object is to solve

$$\min_{\mathbf{x}\in\mathbb{R}^N} \quad \sum_i ||D_i\mathbf{x}||_2 + \tau_1||\varphi\mathbf{x}||_1 + \frac{\tau_2}{2}||A\mathbf{x}-\mathbf{b}||_2^2 \tag{6.14}$$

where $\mathbf{x}$ is the object to be reconstructed, $\sum_i ||D_i\mathbf{x}||_2$ denotes the isotropic total variation norm, φ stands for the othonormal basis, A and $\mathbf{b}$ are the system matrix and the measurements, and τ_1, $\tau_2 > 0$ are the parameters to balance regularization and data fidelity. Here for any i, D_i is a 2-by-N matrix such that the two entires of $D_i\mathbf{x}$ represent the horizontal and vertical local finite differences of $\mathbf{x}$ at pixel i, and the periodic boundary conditions for $\mathbf{x}$ are applied.

In accordance with the ADM model, the auxiliary variables $\mathbf{w} = [\mathbf{w}_1, \cdots, \mathbf{w}_N]$,

$\mathbf{w}_i \in \mathbb{R}^2$ and $\mathbf{z} \in \mathbb{R}^N$ are introduced to transform (6.14) to:

$$\min_{\mathbf{w},\mathbf{z},\mathbf{x}} \quad \sum_i ||\mathbf{w}_i||_2 + \tau_1||\mathbf{z}||_1 + \frac{\tau_2}{2}|||A\mathbf{x} - \mathbf{b}|||_2^2, \quad \text{s.t.} \quad \mathbf{w}_i = D_i\mathbf{x}, \forall i; \quad \mathbf{z} = \varphi\mathbf{x}. \tag{6.15}$$

To tackle the linear constraints, the augmented Lagrangian function of (6.15) is formulated as below:

$$\begin{aligned} \mathcal{L}_{\mathcal{A}}(\mathbf{w},\mathbf{z},\mathbf{x},\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2) &= \sum_i (||\mathbf{w}_i||_2 - (\boldsymbol{\lambda}_2)_i^T(\mathbf{w}_i - D_i\mathbf{x}) + \frac{\alpha_1}{2}||\mathbf{w}_i - D_i\mathbf{x}||_2^2) \\ &\quad + \sum_i \tau_1|z_i - (\boldsymbol{\lambda}_1)_i(z_i - \varphi_i\mathbf{x}) + \frac{\alpha_2}{2}(z_i - \varphi_i\mathbf{x})|^2 \\ &\quad + \frac{\tau_2}{2}|||A\mathbf{x} - \mathbf{b}|||_2^2, \end{aligned} \tag{6.16}$$

where $\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2$ are the Lagrangian multipliers, and $\forall i, (\boldsymbol{\lambda}_1)_i \in \mathbb{R}, (\boldsymbol{\lambda}_2)_i \in \mathbb{R}^2$. φ_i is the ith row of φ. Without generality, it can be assumed that $\alpha_1 = \alpha_2 = \alpha$ for simplicity.

For fixed $\mathbf{x}^t$ and $\boldsymbol{\lambda}^t = (\boldsymbol{\lambda}_1^t, \boldsymbol{\lambda}_2^t)$, the minimizer z_i^{t+1} is given by:

$$\begin{aligned} z_i^{t+1} &= \text{Shrink1}(\varphi_i\mathbf{x}^t + (\boldsymbol{\lambda}_1^t)_i/\alpha, \tau_1/\alpha) \\ &:= \max\{|\varphi_i\mathbf{x}^t + (\boldsymbol{\lambda}_1^t)_i/\alpha| - \tau_1/\alpha, 0\} \circ \text{sgn}(\varphi_i\mathbf{x}^t + (\boldsymbol{\lambda}_1^t)_i/\alpha), \forall i, \end{aligned}$$

and the minimizer $\mathbf{w}_i^{t+1}$ is given by:

$$\begin{aligned} \mathbf{w}_i^{t+1} &= \text{Shrink2}(D_i\mathbf{x}^t + (\boldsymbol{\lambda}_2^t)_i/\alpha, 1/\alpha) \\ &:= \max\{||D_i\mathbf{x}^t + (\boldsymbol{\lambda}_2^t)_i/\alpha||_2 - 1/\alpha, 0\} \cdot \frac{D_i\mathbf{x}^t + (\boldsymbol{\lambda}_2^t)_i/\alpha}{||D_i\mathbf{x}^t + (\boldsymbol{\lambda}_2^t)_i/\alpha||_2}, \forall i. \end{aligned}$$

Let $D^{(1)}$ and $D^{(2)}$ denote the horizontal and vertical global finite difference matrices, $D := (D^{(1)}; D^{(2)})$ and $\mathbf{W}_j := (\mathbf{w}_1(j); \cdots; \mathbf{w}_N(j)), j = 1, 2, \mathbf{W} := (\mathbf{W}_1; \mathbf{W}_2)$. Then the minimization of $\mathcal{L}_{\mathcal{A}}$ can be rewritten as:

$$\begin{aligned} \min_{\mathbf{x}} \quad & -\boldsymbol{\lambda}_2^T(\mathbf{W} - D\mathbf{x}) + \frac{\alpha_1}{2}||\mathbf{W} - D\mathbf{x}||_2^2) \\ & -\boldsymbol{\lambda}_1^T(\mathbf{z} - \varphi\mathbf{x}) + \frac{\alpha_2}{2}||\mathbf{z} - \varphi\mathbf{x}||_2^2 + \frac{\tau_2}{2}|||A\mathbf{x} - \mathbf{b}|||_2^2, \end{aligned}$$

which is equivalent to the normal equation below:

$$M\mathbf{x} = \mathbf{y}, \tag{6.17}$$

where

$$M = D^T D + \varphi^T \varphi + \frac{\tau_2}{\alpha} A^T A,$$

and

$$\mathbf{y} = D^T(\mathbf{W} - \boldsymbol{\lambda}_2/\alpha) + \varphi^T(\mathbf{z} - \boldsymbol{\lambda}_1/\alpha) + (\tau_2/\alpha)A^T\mathbf{b}.$$

Since φ is the othogonal basis, we have $\varphi^T\varphi = I$. It is also noticed that the finite difference matrices $D^{(1)}$ and $D^{(2)}$ are block circulant matrices and can be diagonalized by the discrete Fourier transform F and its inverse F^*, i.e. $\hat{D}^{(j)} = FD^{(j)}F^*, j = 1, 2$ is diagonal. It is known that the system matrix A is a partial Fourier matrix, noted as F_p. Taking these elements into account and for the purpose of simplifying the problem in (6.17), we multiply it by F on both sides.

$$\begin{aligned}
\text{right side} &= F(D^T(\mathbf{W} - \boldsymbol{\lambda}_2/\alpha) + \varphi^T(\mathbf{z} - \boldsymbol{\lambda}_1/\alpha)) + \frac{\tau_2}{\alpha} \cdot P^T\mathbf{b} \\
&= \hat{\mathbf{y}} \\
\text{left side} &= (FD^T DF^* + FIF^* + \frac{\tau_2}{\alpha} FF_p^T F_p F^*) \cdot F\mathbf{x} \\
&= (F({D^{(1)}}^T, {D^{(2)}}^T)F^* F(D^{(1)}; D^{(2)})F^* + I + \frac{\tau_2}{\alpha} P^T P) \cdot F\mathbf{x} \\
&= (\hat{D}^T\hat{D} + I + \frac{\tau_2}{\alpha} P^T P) \cdot F\mathbf{x} \\
&= \hat{M}\hat{\mathbf{x}}
\end{aligned}$$

where $\hat{D} = (\hat{D}^{(1)}; \hat{D}^{(2)}), P = F_p F^*, \hat{\mathbf{x}}$ is the 1D DFT of $\mathbf{x}$, and $\hat{M} = \hat{D}^T\hat{D} + I + \tau_2/\alpha P^T P$ is a diagonal matrix, which makes $\hat{M}\hat{\mathbf{x}} = \hat{\mathbf{y}}$ straightforward to solve. $\mathbf{x}$ is simply the 1-D inverse Fourier transform of $\hat{\mathbf{x}}$.

Finally, the multipliers are updated by:

$$\begin{aligned}
(\boldsymbol{\lambda}_1)_i^{t+1} &= (\boldsymbol{\lambda}_1)_i^t - \eta\alpha(z_i^{t+1} - \varphi_i \mathbf{x}^{t+1}), \forall i \\
(\boldsymbol{\lambda}_2)_i^{t+1} &= (\boldsymbol{\lambda}_2)_i^t - \eta\alpha(\mathbf{w}_i^{t+1} - D_i \mathbf{x}^{t+1}), \forall i
\end{aligned}$$

To sum up, (6.14) is solved via ADM with iterations as below:

$$\begin{cases} z_i^{t+1} & = \text{Shrink1}(\varphi_i \mathbf{x}^t + (\boldsymbol{\lambda}_1^t)_i/\alpha, \tau_1/\alpha), \forall i \\ \mathbf{w}_i^{t+1} & = \text{Shrink2}(D_i \mathbf{x}^t + (\boldsymbol{\lambda}_2^t)_i/\alpha, 1/\alpha), \forall i \\ \mathbf{x}^{t+1} & = F^*(\hat{\mathbf{y}}/\hat{M}) \\ (\boldsymbol{\lambda}_1)_i^{t+1} & = (\boldsymbol{\lambda}_1)_i^t - \eta\alpha(z_i^{t+1} - \varphi_i \mathbf{x}^{t+1}), \forall i \\ (\boldsymbol{\lambda}_2)_i^{t+1} & = (\boldsymbol{\lambda}_2)_i^t - \eta\alpha(\mathbf{w}_i^{t+1} - D_i \mathbf{x}^{t+1}), \forall i \end{cases}$$

where $\hat{\mathbf{y}}$ depends on $(\mathbf{w}^{t+1}, \mathbf{z}^{t+1})$ and $(\boldsymbol{\lambda}_1^t, \boldsymbol{\lambda}_2^t)$. The iterations are terminated simply when the relative change in $\mathbf{x}$ becomes sufficiently small.

For convergence analysis, this algorithm falls into the category of exact ADM. Classic results in the literature [54,55] show that for any $\alpha > 0$ and $\eta \in (0, (\sqrt{5}+1)/2)$, the sequence $\{\mathbf{w}^t, \mathbf{z}^t, \mathbf{x}^t\}$ generated by this algorithm from any starting point $\{\mathbf{w}^0, \mathbf{z}^0, \mathbf{x}^0\}$ converges to a solution of (6.15).

6.3 Experimental Results and Discussion

6.3.1 Noise-free Reconstruction

In this section, two groups of experiments are described over the same Farey angle sets and different Farey angle sets, respectively. The performance of different methods is measured through the quantities listed below: the error in the reconstructed image relative to the original image (Err), signal-to-noise ratio (SNR) and mean square error (MSE). Suppose that f_0 denotes the original image, and f is the reconstructed image, then the definitions of the three quantities are:

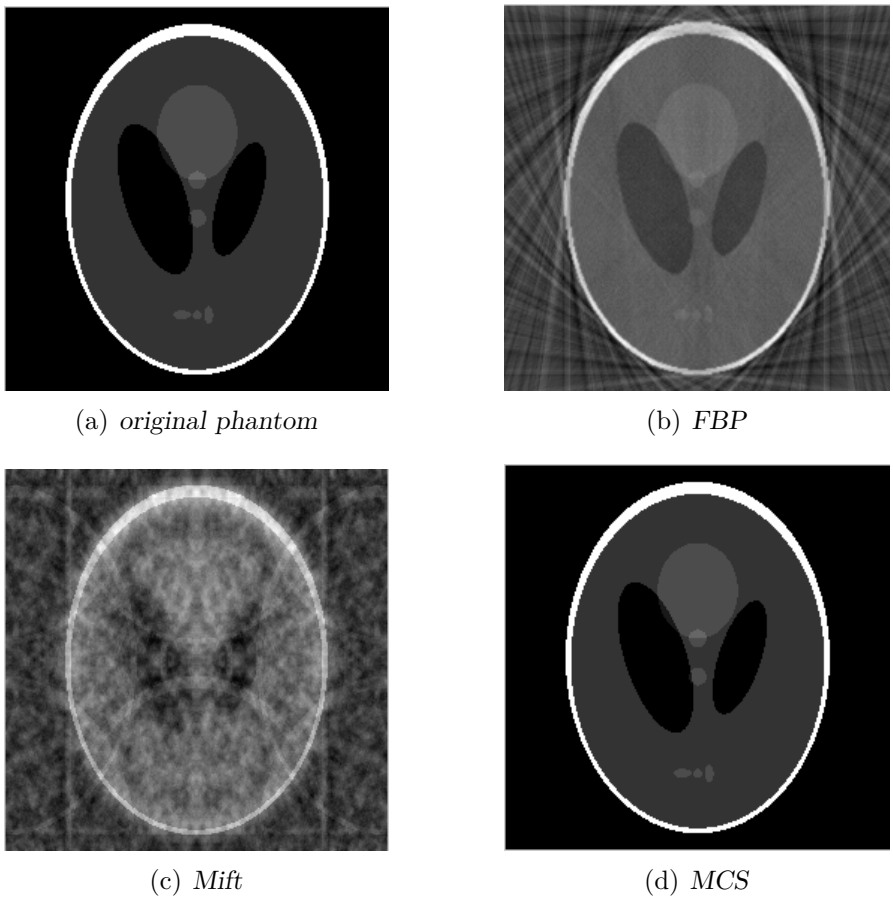
$$\text{Err} = \sqrt{\frac{\sum_i \sum_j (f(i,j) - f_0(i,j))^2}{\sum_i \sum_j f_0(i,j)^2}}$$

$$\text{SNR} = 10 * \lg \frac{\sum_i \sum_j (f(i,j) - \text{mean}(f))^2}{\sum_i \sum_j (f(i,j) - f_0(i,j))^2}$$

$$\text{MSE} = \text{mean}((f(i,j) - f_0(i,j))^2)$$

First, the experiments were conducted over the same Farey angle sets. We present simulation results of the 256×256 phantom image reconstruction obtained by

MCS and two other methods: FBP and Mift, where FBP used filtered back-projection algorithm over the Radon projections and Mift applied direct inverse Fourier transform to the Mojette projections. The projections were taken over the Farey angles. Figure 6.5 shows the reconstruction results of the three methods for F_8 projections, from which we can see that the FBP result has disturbing artifacts due to limited projections, and Mift cannot recover the image from insufficient Fourier samples. It is obvious that MCS produces an almost exact construction. Table 6.1 gives the performance of the above three methods for different orders of Farey series according to Err and SNR. In order to visualize the numerical results intuitive, we plotted the Err of the three different methods as a function of the order of Farey series, and the results are shown in Figure 6.6. We can see that MCS achieves good reconstruction results with F_4 projections, where the Err of Mift and FBP is still larger than 50%. Figure 6.6 also shows that FBP can gain better results than Mift with an increasing number of projections.



(a) *original phantom* (b) *FBP*

(c) *Mift* (d) *MCS*

Figure 6.5: Reconstruction results of the three methods from the projections taken over F_8

Table 6.1: Comparisons of different methods over the same angle set

F	FBP		Mift		MCS	
	Err	SNR	Err	SNR	Err	SNR
F1	235.56%	-8.7	73.81%	1.4	75.2%	1.2
F2	147.72%	-4.6	70.25%	1.8	63.08%	2.8
F3	92.05%	-0.5	65.45%	2.4	13.57%	16.1
F4	71.08%	1.7	62.87 %	2.8	0.57%	43.7
F5	51.36%	4.5	60.46%	3.1	0.16%	54.7
F6	47.24%	5.3	59.35%	3.3	0.1%	58.4
F7	39.58%	6.8	53.33%	4.2	0.01%	78.1
F8	37.16%	7.4	50.82%	4.6	0.00%	89.4

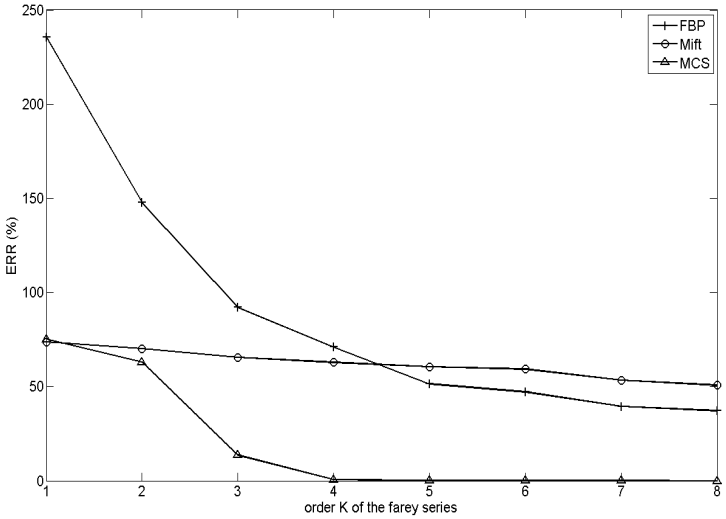


Figure 6.6: A comparison of Err gained by different methods with the increasing order of the Farey series

Next the experiments were conducted over different Farey angle sets. The Mojette filtered backprojection algorithm (MFBP) [134] is used here for comparison. We used the same experimental image as MFBP, a 128×128 phantom image consisting of a 17×17 square object with unitary value, whereas the boundaries were only half valued. Table 6.2 lists the MSE of MCS reconstruction results with different orders of Farey series, while Table 6.3 is with MFBP. We can see that MCS can gain better results over F_1 angle set with just 4 projection lines than MFBP over F_{32} angle set with 1296 projection lines. The reconstruction

results are shown in Figure 6.7. The result of MFBP over F_{32} angle in (b) has some disturbing artifacts, while MCS gives clear and exact reconstruction results with only 4 projection lines. Figure 6.7 (d) shows the line mask of the partial Fourier coefficients which are the input of CS. If the X-ray tube current is fixed, the total radiation exposure of the target is proportional to the number of view angles. Hence, the conclusion is drawn that fewer projections are needed for the reconstruction with the help of CS, which means shorter scanning time and lower radiation dosage, and thus will benefit patients more.

Table 6.2: MSE of MCS reconstruction results

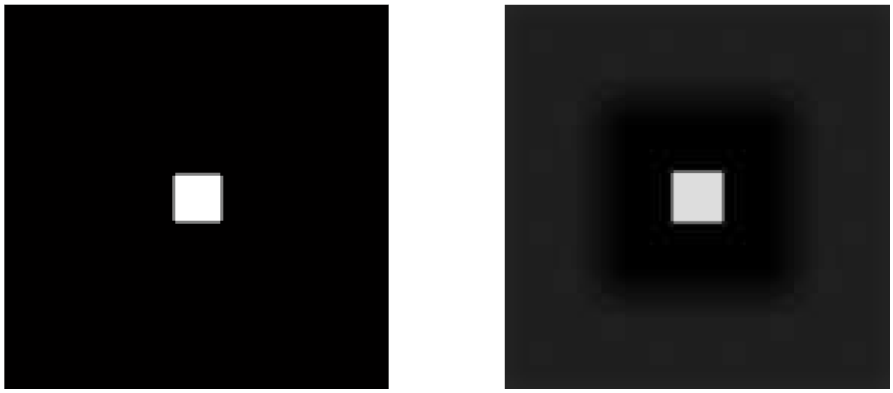
F	F1	F2	F3	F4
#proj	4	8	16	24
MSE	1.69×10^{-7}	6.95×10^{-8}	7.88×10^{-9}	1.23×10^{-10}

Table 6.3: MSE of MFBP reconstruction results

F	F32	F64	F128
#proj	1296	5040	20088
MSE	0.01322	2×10^{-5}	0

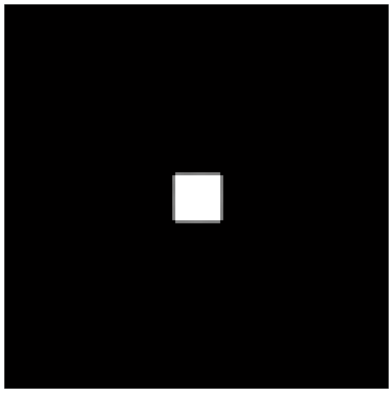
6.3.2 Noise Tolerance

This experiment was designed to test the noise tolerance of the proposed frame. We generated our test sets using the 256×256 Shepp-Logan phantom image and FBP was employed for comparison. The Gaussian noise was added to the Mojette projections in MCS and the Radon projections in FBP, respectively. In both methods, the projections were taken over F_8 angle set, i.e., 88 projections. As the noise was produced randomly, the quantitative assessment (Err and MSE) was the mean value of the 20 groups of experimental data, and the results are shown in Table 6.4. We can see that for the Gaussian noise $(0, 0.001)$, i.e. with mean 0 and standard deviation 0.001, MCS achieves a much better reconstruction than FBP, the results of which are shown in Figure 6.8. But for $(0, 0.01)$ noise, the result of FBP barely changes, while the reconstruction quality of MCS reduces greatly, which means FBP is robust to noise and MCS is sensitive. The reason is that MCS is based on noise-sensitive Fourier transform. When (0,0.01) noise

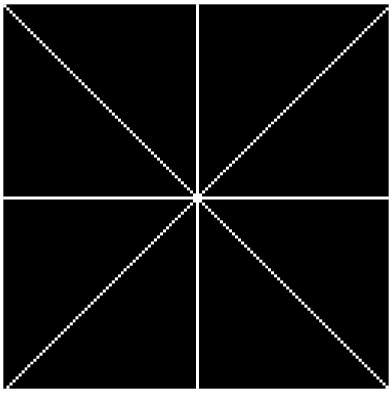


(a) *original phantom*

(b) *MFBP over* F_{32}



(c) *MCS over* F_1



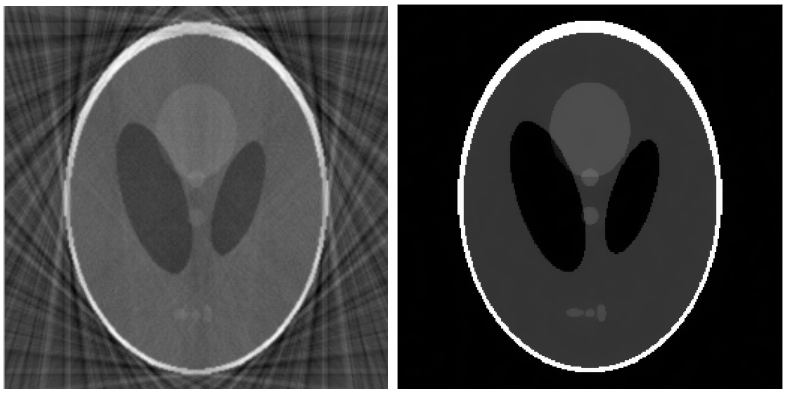
(d) *line mask of* F_1

Figure 6.7: Reconstruction results of MCS and MFBP

is added to the projections, the noise of its Fourier coefficients becomes (0,0.1) for a 256×256 image, as verified in Figure 6.9. Since CS has some ability to suppress noise, we conclude that though sensitive, the proposed frame can deal with small noise effectively. It will be the subject of future work to deal with noisier projections [126].

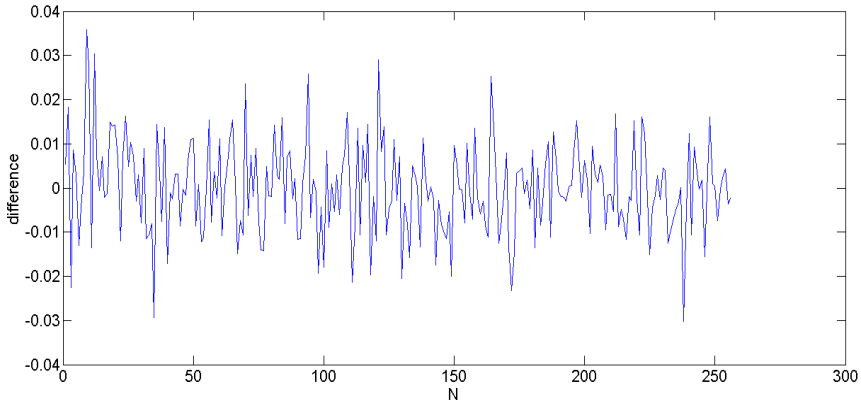
Table 6.4: Noise response of FBP and MCS

Noise	FBP		MCS	
	Err	SNR	Err	SNR
(0,0.001)	37.1577%	7.3586	2.1929%	31.9393
(0,0.01)	37.1587%	7.3584	23.1656%	11.4631

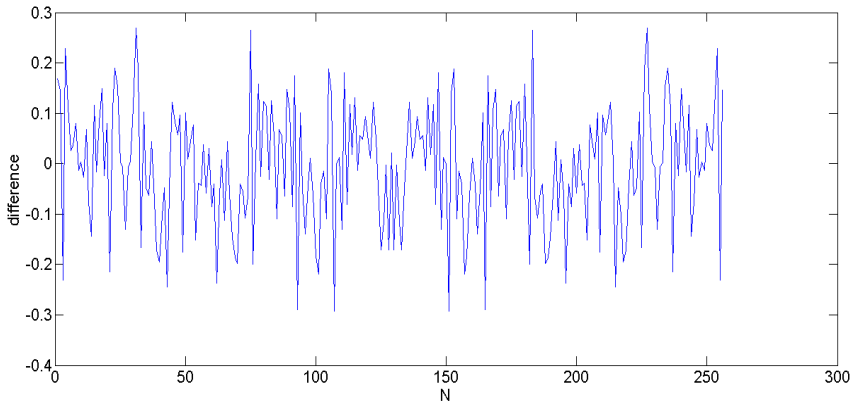


(a) *FBP* (b) *MCS*

Figure 6.8: Reconstruction results from the F_8 projections with Gaussian noise $(0, 0.001)$



(a) *Noise of projections*



(b) *Corresponding noise of the Fourier values*

Figure 6.9: Noise of the projections and the Fourier values

6.4 Conclusion

In this chapter, we have gained more understanding of the direct application of CS in CT reconstruction by presenting a novel frame for parallel-beam CT. First, the sinogram was converted to the projections gained through Mojette transform, which is an exact discretization of the Radon transform. On each view angle, the projections were summed up with the application of certain principles. Then the 1-D Fourier coefficients of the merged projections were mapped to the 2-D Fourier domain of the object. Finally, compressed sensing was employed to deal with the partial Fourier coefficients via the specific optimization algorithm of RecPF, which recovered the object very well and suppressed the small noise effectively. Experimental results have demonstrated the advantages of the proposed method. Using Mojette transform and compressed sensing, the purpose of reducing the radiation dosage during CT examinations without compromising the image quality is achieved.

Chapter 7

Compressed Sensing-Based Sinogram Inpainting for CT Reconstruction

7.1 Introduction

As stated earlier, increased X-ray radiation can lead to ionization of body cells and raise the risk of cancer. Therefore to reduce the X-ray radiation dose without compromising imaging quality has been an important and challenging problem in CT development. The recent technique called compressed sensing (CS) [21–23, 40, 42, 95, 149] has shown great potential. Unlike the conventional reconstruction method, e.g. filtered back-projection (FBP), it does not require a large number of X-ray radiation projections for satisfactory image formation and reconstruction according to the Shannon-Nyquist sampling theorem. Rather, it is efficient and powerful for recovering signals from incomplete measurements and has led to applications in low radiation CT image reconstruction [29, 61, 68, 70, 139, 140].

Recent studies on CT reconstruction by CS have been on the computational efficiency [110, 116] and the improvement of reconstruction accuracy. Typically, the work reported in [14] exploits the sparsity of objects in the total variation (TV) domain, which is denoted as the CSTV model. In [88] and [89], the a priori information is introduced as constraints of CS reconstruction in conjunction with

TV minimization. There are generally two difficulties in the existing applications of CS to CT reconstruction. The first is that as most CT images may not be piecewise smooth, they are not sparse in the TV domain. The second is that the fundamental CS theory requires that the compressed sampling scheme has an incoherence property, in terms of the restricted isometry property (RIP) of the CS sensing matrix, for stable reconstruction. To date there have been few studies of CS reconstruction of CT images under the RIP condition. For example, there has been a study of the analysis of RIP [73] and an empirical study of the incoherence property in terms of the phase transition map [74].

This chapter addresses the above problems to exploit the formulation of a sensing matrix for CT data acquisition which satisfies the RIP. In CT scanning and data acquisition, the physical CT scan performs a mathematical Radon transform. The set of the scanned measurement data is the sinogram, in which each measurement datum represents the intensity of the attenuated X-ray radiation along a path through the patients' body. Under normal CT scan conditions, it is observed that the sinogram is a smooth function of the X-ray radiation angle. Correspondingly, its frequency domain representation is narrow band and hence sparse. Based on this observation, we consider here the transformed sinogram in the frequency domain as the reconstructed signal, which can satisfy the sparsity condition. If the CT scan randomly collects under-sampled sinogram data, it leads to a randomly sampled partial Fourier matrix as the sensing matrix for the CT reconstruction, which is known to satisfy the RIP. As a result, the RIP and sparsity conditions required by CS are satisfied for the CT reconstruction. Once the transformed sinogram in the frequency domain is reconstructed, it can be further processed to produce the CT image. The under-sampled sinogram data set enables considerably reduced X-ray projection operations and hence reduced X-ray radiation dosage to the patient.

The rest of the chapter is organized as follows. Section 7.2 presents brief formulations of fan-beam CT data acquisition and the relation between fan-beam and parallel-beam projections. The proposed CS sampling scheme and the regularized optimization algorithm for CT reconstruction are presented in Section 7.3. Section 7.4 shows the experimental results to demonstrate the advantages of the proposed CS approach to CT reconstruction. Performance comparisons are shown between the proposed weighted ℓ_1 regularized optimization ($w\ell_1$) algorithm and the existing CSTV-algorithm in terms of the visual effect and the reconstruction

error. Conclusions are drawn in Section 7.5.

7.2 Formulation of Fan-beam CT Data Acquisition

The main part of a typical CT scanner is a doughnut-shaped gantry that has a set of X-ray sources and detectors on opposite sides. The sources emit X-rays that are attenuated when radiating through the object inside the gantry. The attenuated X-ray intensities are projected on the detectors to form a set of measurement data, from which CT images representing attenuation coefficients of the scanned object can be reconstructed. Let the scanning object be $f(x, y)$ in a Cartesian coordinate system (x, y). The CT scan maps the object into a set of sinogram data given by

$$s(\theta, r) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\delta(r - x\cos\theta - y\sin\theta)dxdy \tag{7.1}$$

where θ and r are respectively the angular and radial coordinates of the projection line (θ, r), and $\delta(\cdot)$ is the Dirac delta function. Without loss of generality, the sinogram data sets considered in the rest of this chapter are sampled discrete data sets.

The fan-beam data acquisition geometry has been described in Section 5.2.1. Suppose that the fan-beam data set is denoted by $s_f(\beta, \gamma)$. The relation between the independent variables of the fan-beam projections and parallel projections is

$$r = R\sin\gamma \quad \text{and} \quad \theta = \beta + \gamma. \tag{7.2}$$

It follows that the corresponding sinogram data set $s(\theta, r)$ is

$$s(\theta, r) = s_f(\theta - \sin^{-1}\frac{r}{R}, \sin^{-1}\frac{r}{R}). \tag{7.3}$$

For any fixed r, the sinogram $s(\theta, r)$ satisfies $s(0, r) = s(\pi, -r)$ and $s(\pi, r) = s(0, -r)$. Thus $s(\theta, r)$ is a continuous function of θ over $\theta \in [0, 2\pi)$ if it is continuous over $\theta \in [0, \pi)$.

7.3 The CS Approach to CT Image Reconstruction

7.3.1 Formulation of the CS Problem

It is known that the CS method can achieve signal recovery from much fewer measurements than those usually required by the Shannon-Nyquist criterion. Physically and practically, the objective of CT data acquisition with low X-ray radiation directly requires reduced sampled numbers of the sinogram data set $s(\theta, r)$. Theoretically, the CS reconstruction of the CT image further requires that the sampling scheme of the sinogram data set $s(\theta, r)$ meets the RIP and there is a wavelet domain in which the transformed CT image is sparse or compressible [21–23].

To meet the above conditions for the CS reconstruction of CT images, a basic consideration is that, under normal CT scan conditions and for any value of r, the sinogram $s(\theta, r)$ is a continuous function with smooth values of the radiation angle θ for $\theta \in [0, 2\pi)$. For any fixed $r \geq 0$, let $\mathbf{s}_r$, with appropriate dimension, be the vectorized representation of the sinogram $s(\theta, r)$ for $\theta \in [0, 2\pi)$ and $\hat{\mathbf{s}}_r$ be the discrete Fourier transform (DFT) of $\mathbf{s}_r$. There exist DFT matrices F for all r such that

$$\hat{\mathbf{s}}_r = F\mathbf{s}_r. \tag{7.4}$$

Correspondingly, the measurement sinogram data vector is written as

$$\mathbf{s}_r = F^*\hat{\mathbf{s}}_r. \tag{7.5}$$

For the sinogram $s(\theta, r)$, being a continuous function of θ with smooth values, its DFT data vector $\hat{\mathbf{s}}_r$ is narrow band and hence sparse in the frequency domain. If this data vector is considered to be the reconstructed signal, it can meet the sparse or compressible condition for the CS reconstruction.

In practical CT scans, the rotational CT scan mechanism can be controlled to radiate fan-beam X-rays at randomly located angular positions. The randomly controlled angular radiation and data acquisition positions in a range of β can directly result in a set of under-sampled sinogram data vectors denoted by $\tilde{\mathbf{s}}_r$, which is composed of randomly selected entries of $\mathbf{s}_r$. Corresponding to the

vector $\tilde{\mathbf{s}}_r$ of the random entries, there is a partial Fourier matrix of appropriate dimension, denoted by $\mathbf{\Phi}_r$. It is composed of rows of the inverse DFT matrix F^* corresponding to the random entries of $\tilde{\mathbf{s}}_r$, such that

$$\tilde{\mathbf{s}}_r = \mathbf{\Phi}_r \hat{\mathbf{s}}_r. \tag{7.6}$$

It follows from the known CS theory [21–23] that the partial Fourier matrix $\mathbf{\Phi}_r$ satisfies the RIP.

As a result, the formulation of the CS reconstruction of fan-beam CT images is given by (7.6). It is based on the randomly under-sampled sinogram data vector $\tilde{\mathbf{s}}_r$ for reconstructing the sparse DFT data vector $\hat{\mathbf{s}}_r$. Since the corresponding sensing matrix $\mathbf{\Phi}_r$ is a randomly under-sampled partial Fourier matrix, it satisfies the RIP. The CS reconstructed $\hat{\mathbf{s}}_r$, as the DFT of the complete sinogram data vector, will contain sufficient information for further reconstruction of the CT image.

7.3.2 The CS Reconstruction Frame

Given the under-sampled sinogram data vector $\tilde{\mathbf{s}}_r$ and the CS reconstruction formulation in the form (7.6), a number of algorithms are available for computation of the reconstruction solution. The popular basis pursuit algorithm is formulated as

$$\min \|\hat{\mathbf{s}}_r\|_1, \text{ subject to } \tilde{\mathbf{s}}_r = \mathbf{\Phi}_r \hat{\mathbf{s}}_r, \quad \forall r, \tag{7.7}$$

where $\|\cdot\|_1$ denotes the ℓ_1 norm. The problem can be alternatively formulated as

$$\min_{\hat{\mathbf{s}}_r} \frac{1}{2\sigma} \|\tilde{\mathbf{s}}_r - \mathbf{\Phi}_r \hat{\mathbf{s}}_r\|_2^2 + \|\hat{\mathbf{s}}_r\|_1, \quad \forall r, \tag{7.8}$$

where σ is a weighting parameter.

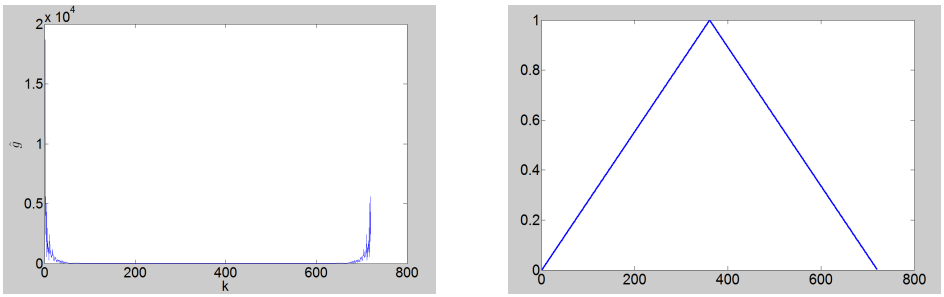
To present a more relevant and efficient CS algorithm for the considered CT reconstruction problem, the narrow frequency band property of the frequency domain data vector $\hat{\mathbf{s}}_r$ is taken into account. Following from the result of [25] on the weighted CS technique, weighting factors w_i are introduced to define the

weighted ℓ_1 ($w\ell_1$) norm of $\hat{\mathbf{s}}_r$ as

$$\|\hat{\mathbf{s}}_r\|_{w1} = \sum_i |w_i \hat{s}_{ri}|, \tag{7.9}$$

where $\hat{s}_{ri}$ denotes the ith entry of $\hat{\mathbf{s}}_r$. The guideline for the selection of the weighting factors w_i is that large weighting factors are applied to small entries of the reconstruction vector. Based on the fact that the frequency domain vector $\hat{\mathbf{s}}_r$ is in general of low pass nature, typically with the magnitude frequency plot as shown in Fig. 7.1(a), the weighting factors selected for the weighted ℓ_1 norm and the CS reconstruction algorithm in this chapter is of a triangle shape, as shown in Fig. 7.1(b). Using the weighting factors and the weighted ℓ_1 norm, the weighted ℓ_1 regularized optimization algorithm for the CS reconstruction of CT images is formulated as

$$\min_{\hat{\mathbf{s}}_r} \quad \frac{1}{2\sigma}\|\tilde{\mathbf{s}}_r - \tilde{\mathbf{\Phi}}_r \hat{\mathbf{s}}_r\|_2^2 + \|\hat{\mathbf{s}}_r\|_{w1}, \quad \forall r. \tag{7.10}$$



(a) *Typical magnitude frequency plot* (b) *Plot of weighting factors*

Figure 7.1: Plots of magnitude frequency and weighting factors.

The weighted ℓ_1 regularized optimal CS reconstruction algorithm computes a solution for the DFT vectors $\hat{\mathbf{s}}_r$, $\forall r$. The sinogram vectors $\mathbf{s}_r$, $\forall r$, can be obtained by taking the inverse DFTs of $\hat{\mathbf{s}}_r$. The FBP operation can be further applied to obtain the reconstructed CT image.

7.3.3 Optimization Algorithm – Yall1

The proposed problem in (7.10) is solved by the algorithm called Yall1 [159]. It is a first-order primal-dual algorithm based on the alternating direction method (ADM).

Applying ADM to primal problems

The object is to solve:

$$\min_{\mathbf{x}\in\mathbb{C}^N} \quad ||\mathbf{x}||_1 + \frac{1}{2\sigma}||A\mathbf{x}-\mathbf{b}||_2^2. \tag{7.11}$$

To start with, (7.11) is reformulated in accordance with (5.8) by intoducing an auxiliary variable $\mathbf{y}\in\mathbb{C}^m$:

$$\min_{\mathbf{x}\in\mathbb{C}^N,\mathbf{y}\in\mathbb{C}^m} \quad ||\mathbf{x}||_1 + \frac{1}{2\sigma}||\mathbf{y}||_2^2 \quad \text{s.t.} \quad A\mathbf{x}+\mathbf{y}=\mathbf{b} \tag{7.12}$$

Its augmented Lagrangian function is written as:

$$\min_{\mathbf{x}\in\mathbb{C}^N,\mathbf{y}\in\mathbb{C}^m} \quad ||\mathbf{x}||_1 + \frac{1}{2\sigma}||\mathbf{y}||_2^2 - \mathbf{Re}(\boldsymbol{\lambda}^*(A\mathbf{x}+\mathbf{y}-\mathbf{b})) + \frac{\alpha}{2}||A\mathbf{x}+\mathbf{y}-\mathbf{b}||_2^2, \tag{7.13}$$

where $\boldsymbol{\lambda}\in\mathbb{C}^m$ is the Lagrangian multiplier and $\alpha>0$ is the penalty parameter. Firstly, given $(\mathbf{x}^t,\boldsymbol{\lambda}^t)$, the minimizer of (7.13) with respect to $\mathbf{y}$ is given by:

$$\mathbf{y}^{t+1} = \frac{\sigma\alpha}{\sigma\alpha+1}(\boldsymbol{\lambda}^t/\alpha - (A\mathbf{x}^t-\mathbf{b})). \tag{7.14}$$

Secondly, with fixed $\mathbf{y}^{t+1}$ and $\boldsymbol{\lambda}^t$, the minimizer of (7.13) with respect to $\mathbf{y}$ is equivalent to:

$$\min_{\mathbf{x}\in\mathbb{C}^N} \quad ||\mathbf{x}||_1 + \frac{\alpha}{2}||A\mathbf{x}+\mathbf{y}^{t+1}-\mathbf{b}-\boldsymbol{\lambda}/\alpha||_2^2 \tag{7.15}$$

Defining $\mathbf{r}^t = A^*(A\mathbf{x}^t+\mathbf{y}^{t+1}-\mathbf{b}-\boldsymbol{\lambda}^t/\alpha)$ as the gradient of the quadratic term in (7.15) and $\tau>0$ the proximal parameter, (7.15) is appoximated by:

$$\min_{\mathbf{x}\in\mathbb{C}^N} \quad ||\mathbf{x}||_1 + \alpha((\mathbf{r}^t)^*(\mathbf{x}-\mathbf{x}^t) + \frac{1}{2\tau}||\mathbf{x}-\mathbf{x}^t||_2^2) \tag{7.16}$$

Making use of the well known 1-D shrinkage, or soft thresholding, the solution of (7.16) is given by:

$$\mathbf{x}^{t+1} = \text{Shrink1}(\mathbf{x}^t-\tau\mathbf{r}^t, \frac{\tau}{\alpha}) := \max\{|\mathbf{x}^t-\tau r^t| - \frac{\tau}{\alpha}, 0\} \circ \text{sgn}(\mathbf{x}^t-\tau r^t) \tag{7.17}$$

where $\circ$ denotes component-wise multiplication, and $\text{sgn}(z) := \begin{cases} -1 & \text{if } z<0 \\ 0 & \text{if } z=0 \\ 1 & \text{if } z>0 \end{cases}$.

Finally, the multiplier $\boldsymbol{\lambda}$ is updated by:

$$\boldsymbol{\lambda}^{t+1} = \boldsymbol{\lambda}^t - \eta\alpha(A\mathbf{x}^{t+1} + \mathbf{y}^{t+1} - \mathbf{b}).$$

with the constant $\eta > 0$.

To sum up, (7.11) is solved via ADM with iterations as below:

$$\begin{cases} \mathbf{y}^{t+1} = & \frac{\sigma\alpha}{\sigma\alpha+1}(\boldsymbol{\lambda}^t/\alpha - (A\mathbf{x}^t - \mathbf{b})) \\ \mathbf{x}^{t+1} = & \text{Shrink1}(\mathbf{x}^t - \tau\mathbf{r}^t, \frac{\tau}{\alpha}) \\ \boldsymbol{\lambda}^{t+1} = & \boldsymbol{\lambda}^t - \eta\alpha(A\mathbf{x}^{t+1} + \mathbf{y}^{t+1} - \mathbf{b}) \end{cases} \tag{7.18}$$

The convergence of the primal-based algorithm is established below.

Theorem 7.1. *[159] Let $\tau, \eta > 0$ satisfy $\tau\lambda_{max} + \eta < 2$, where λ_{max} denotes the maximum eigenvalue of A^*A. For any fixed $\alpha > 0$, the sequence $\{(\mathbf{y}^t, \mathbf{x}^t, \boldsymbol{\lambda}^t)\}$ generated by (7.18) from any starting point $(\mathbf{x}^0, \boldsymbol{\lambda}^0)$ converges to $(\tilde{\mathbf{y}}^t, \tilde{\mathbf{x}}^t, \tilde{\boldsymbol{\lambda}}^t)$, where $(\tilde{\mathbf{y}}^t, \tilde{\mathbf{x}}^t)$ is the solution of (7.12).*

The proof is given in Section 7.3.4.

Applying ADM to dual problems

The dual problem of (7.11) is given by:

$$\begin{aligned} & \max_{\boldsymbol{\lambda}\in\mathbb{C}^m} \min_{\mathbf{x}\in\mathbb{C}^N, \mathbf{y}\in\mathbb{C}^m} \{||\mathbf{x}||_1 + \frac{1}{2\sigma}||\mathbf{y}||_2^2 - \mathbf{Re}(\boldsymbol{\lambda}^*(A\mathbf{x} + \mathbf{y} - \mathbf{b}))\} \\ = & \max_{\boldsymbol{\lambda}\in\mathbb{C}^m} \{\mathbf{Re}(\mathbf{b}^*\boldsymbol{\lambda}) - \frac{\sigma}{2}||\boldsymbol{\lambda}||_2^2 + \min_{\mathbf{x}\in\mathbb{C}^N}(||\mathbf{x}||_1 - \mathbf{Re}(\boldsymbol{\lambda}^*A\mathbf{x})) + \frac{1}{2\sigma}\min_{\mathbf{y}\in\mathbb{C}^m}||\mathbf{y} - \sigma\boldsymbol{\lambda}||_2^2\} \\ = & \max_{\boldsymbol{\lambda}\in\mathbb{C}^m} \{\mathbf{Re}(\mathbf{b}^*\boldsymbol{\lambda}) - \frac{\sigma}{2}||\boldsymbol{\lambda}||_2^2\} \quad \text{s.t.} \quad A^*\boldsymbol{\lambda} \in \mathbf{B}_1^\infty \end{aligned} \tag{7.19}$$

where $\mathbf{B}_1^\infty := \{\boldsymbol{\xi} \in \mathbb{C}^N : ||\boldsymbol{\xi}||_\infty \leq 1\}$, $||\cdot||_\infty$ denotes the maximum of $\boldsymbol{\xi}$. The dual problem can be also solved under the frame of ADM, which appears to be more efficient.

Similar to the primal problem, (7.19) is reformulated by introducing an auxiliary

variable $\mathbf{z} \in \mathbb{C}^N$:

$$\max_{\boldsymbol{\lambda} \in \mathbb{C}^m} \quad \mathbf{Re}(\mathbf{b}^* \boldsymbol{\lambda}) - \frac{\sigma}{2} ||\boldsymbol{\lambda}||_2^2 \quad \text{s.t.} \quad \mathbf{z} - A^* \boldsymbol{\lambda} = 0, \quad \mathbf{z} \in \mathbf{B}_1^\infty.$$

Its augmented Lagrangian function is written as:

$$\min_{\boldsymbol{\lambda} \in \mathbb{C}^m, \mathbf{z} \in \mathbb{C}^N} \{ -\mathbf{Re}(\mathbf{b}^* \boldsymbol{\lambda}) + \frac{\sigma}{2} ||\boldsymbol{\lambda}||_2^2 - \mathbf{Re}(\mathbf{x}^* (\mathbf{z} - A^* \boldsymbol{\lambda})) + \frac{\alpha}{2} ||\mathbf{z} - A^* \boldsymbol{\lambda}||_2^2, \mathbf{z} \in \mathbf{B}_1^\infty \}$$

where $\mathbf{x} \in \mathbb{C}^N$ is a multiplier as well as the primal variable and α is the penalty parameter. Given $(\mathbf{x}^t, \boldsymbol{\lambda}^t)$, the minimizer $\mathbf{z}^{t+1}$ is given by:

$$\mathbf{z}^{t+1} = \mathcal{Q}_{\mathbf{B}_1^\infty}(A^* \boldsymbol{\lambda}^t + \mathbf{x}^t / \alpha),$$

where $\mathcal{Q}_{\mathbf{B}_1^\infty}$ stands for the projection in Euclidean norm onto the convex set $\mathbf{B}_1^\infty$. Then, with fixed $\mathbf{x}^t$ amd $\mathbf{z}^{t+1}$, the minimizer $\boldsymbol{\lambda}$ is:

$$\boldsymbol{\lambda}^{t+1} = \frac{\alpha}{\alpha A A^* + \sigma} (A \mathbf{z}^{t+1} - (A \mathbf{x}^t - \mathbf{b}) / \alpha).$$

Finally, $\mathbf{x}$ is updated by:

$$\mathbf{x}^{t+1} = \mathbf{x}^t - \eta \alpha (\mathbf{z}^{t+1} - A^* \boldsymbol{\lambda}^{t+1}),$$

with constant $\eta \in (0, (\sqrt{5} - 1)/2)$.

In conclusion, the dual problem of (7.11) is solved via ADM with iterations as below:

$$\begin{cases} \mathbf{z}^{t+1} = & \mathcal{Q}_{\mathbf{B}_1^\infty}(A^* \boldsymbol{\lambda}^t + \mathbf{x}^t / \alpha) \\ \boldsymbol{\lambda}^{t+1} = & \frac{\alpha}{\alpha A A^* + \sigma} (A \mathbf{z}^{t+1} - (A \mathbf{x}^t - \mathbf{b}) / \alpha) \\ \mathbf{x}^{t+1} = & \mathbf{x}^t - \eta \alpha (\mathbf{z}^{t+1} - A^* \boldsymbol{\lambda}^{t+1}) \end{cases}$$

There are two cases for the update of $\boldsymbol{\lambda}$. One is when $AA^* = I$ and the update is simple and straightforward, written as $\boldsymbol{\lambda}^{t+1} = \frac{\alpha}{\alpha + \sigma} (A \mathbf{z}^{t+1} - (A \mathbf{x}^t - \mathbf{b}) / \alpha)$. The other case is when $AA^* \neq I$ and the update is costly. In this case, a steepest descent step in $\boldsymbol{\lambda}$ direction is adopted:

$$\boldsymbol{\lambda}^{t+1} = \boldsymbol{\lambda}^t - \varsigma_t^* \mathbf{g}^t,$$

where $\mathbf{g}^t = \sigma \boldsymbol{\lambda}^t + A \mathbf{x}^t - \mathbf{b} + \alpha A (A^* \boldsymbol{\lambda}^t - \mathbf{z}^{t+1})$ and $\varsigma_t^* = \frac{(\mathbf{g}^t)^* \mathbf{g}^t}{(\mathbf{g}^t)^* (\sigma I + \alpha A A^*) \mathbf{g}^t}$.

The convergence of the dual-based algorithm is also analyzed [159]. Under the assumption $A^*A = I$, the algorithm falls into the category of exact ADM. Results reported in the literature [54,55] show that for any $\alpha > 0$ and $\eta \in (0, (\sqrt{5}+1)/2)$, the sequence $\{\mathbf{z}^t, \boldsymbol{\lambda}^t, \mathbf{x}^t\}$ generated by the dual-based algorithm from any starting point $(\mathbf{x}^0, \boldsymbol{\lambda}^0)$ converges to $(\tilde{\mathbf{z}}^t, \tilde{\boldsymbol{\lambda}}^t, \tilde{\mathbf{x}}^t)$, which is the solution of the primal-dual pair (7.11) and (7.19). The convergence remains an issue of further research for the case when $A^*A \neq I$, however, it is shown experimentally to converge very well for random matrix A.

7.3.4 Proof of Theorem 7.1

Let $(\tilde{\mathbf{y}}, \tilde{\mathbf{x}})$ be any solution of (7.12). The optimization theory states that there exists $\tilde{\boldsymbol{\lambda}} \in \mathbb{C}^m$ such that:

$$\tilde{\mathbf{y}}/\sigma - \tilde{\boldsymbol{\lambda}} = \mathbf{0}, \quad A^*\tilde{\mathbf{y}} \in \partial||\tilde{\mathbf{x}}||_1 \quad \text{and} \quad A\tilde{\mathbf{x}} + \tilde{\mathbf{y}} = \mathbf{b}. \tag{7.20}$$

Suppose

$$\hat{\boldsymbol{\lambda}} := \boldsymbol{\lambda}^t - \alpha(A\mathbf{x}^{t+1} + \mathbf{y}^{t+1} - \mathbf{b}) \tag{7.21}$$

then $\boldsymbol{\lambda}^{t+1} = \boldsymbol{\lambda}^t - \eta(\boldsymbol{\lambda}^t - \hat{\boldsymbol{\lambda}})$.

The update of $\mathbf{y}$ in (7.14) can be reformulated as $\mathbf{y}^{t+1}/\delta - \hat{\boldsymbol{\lambda}} + \alpha A(\mathbf{x}^t - \mathbf{x}^{t+1}) = 0$. Considering the first equation in (7.20), we have $\hat{\boldsymbol{\lambda}} - \tilde{\boldsymbol{\lambda}} - \alpha A(\mathbf{x}^t - \mathbf{x}^{t+1}) = (\mathbf{y}^{t+1} - \tilde{\mathbf{y}})/\sigma$, and thus

$$(\mathbf{y}^{t+1} - \tilde{\mathbf{y}})^*(\hat{\boldsymbol{\lambda}} - \tilde{\boldsymbol{\lambda}} - \alpha A(\mathbf{x}^t - \mathbf{x}^{t+1})) = ||\mathbf{y}^{t+1} - \tilde{\mathbf{y}}||_2^2/\sigma \geq 0. \tag{7.22}$$

In the yall1 algorithm, the update of $\mathbf{x}$ is not updated exactly, and instead we use (7.16) to approximate (7.15). Hence, it is necessary to conduct the convergence analysis. (7.16) implies:

$$\begin{aligned} \alpha A^*(A\mathbf{x}^t + \mathbf{y}^{t+1} - \mathbf{b} - \boldsymbol{\lambda}^t/\alpha) + \frac{\alpha}{\tau}(\mathbf{x}^{t+1} - \mathbf{x}^t) &\in \partial||\mathbf{x}^{t+1}||_1 \\ \text{(from (7.21))} \quad A^*\hat{\boldsymbol{\lambda}} - \alpha A^*A(\mathbf{x}^t - \mathbf{x}^{t+1}) + \frac{\alpha}{\tau}(\mathbf{x}^t - \mathbf{x}^{t+1}) &\in \partial||\mathbf{x}^{t+1}||_1 \end{aligned}$$

Considering the second equation in (7.20) and the convexity of $||\cdot||_1$, we have:

$$(\mathbf{x}^{t+1}-\tilde{\mathbf{x}})^*(A^*(\hat{\boldsymbol{\lambda}}-\tilde{\boldsymbol{\lambda}})-\alpha A^*A(\mathbf{x}^t-\mathbf{x}^{t+1})+\frac{\alpha}{\tau}(\mathbf{x}^t-\mathbf{x}^{t+1}))\geq 0$$

It is known from (7.20) and (7.21) that $(\mathbf{x}^{t+1}-\tilde{\mathbf{x}})^*A^*=\frac{1}{\alpha}(\boldsymbol{\lambda}^t-\hat{\boldsymbol{\lambda}})^*-(\mathbf{y}^{t+1}-\tilde{\mathbf{y}})^*$. If we substitute it into the inequality, we obtain:

$$\begin{aligned}
&\frac{1}{\alpha}(\hat{\boldsymbol{\lambda}}-\tilde{\boldsymbol{\lambda}})^*(\boldsymbol{\lambda}^t-\hat{\boldsymbol{\lambda}})+\frac{\alpha}{\tau}(\mathbf{x}^{t+1}-\tilde{\mathbf{x}})^*(\mathbf{x}^t-\mathbf{x}^{t+1})\\
&-(\boldsymbol{\lambda}^t-\hat{\boldsymbol{\lambda}})A(\mathbf{x}^t-\mathbf{x}^{t+1})-(\mathbf{y}^{t+1}-\tilde{\mathbf{y}})^*(\hat{\boldsymbol{\lambda}}-\tilde{\boldsymbol{\lambda}}-\alpha A(\mathbf{x}^t-\mathbf{x}^{t+1}))\geq 0\\
&\text{(from (7.22))}\\
&\frac{1}{\alpha}(\hat{\boldsymbol{\lambda}}-\tilde{\boldsymbol{\lambda}})^*(\boldsymbol{\lambda}^t-\hat{\boldsymbol{\lambda}})+\frac{\alpha}{\tau}(\mathbf{x}^{t+1}-\tilde{\mathbf{x}})^*(\mathbf{x}^t-\mathbf{x}^{t+1})\\
&\geq(\boldsymbol{\lambda}^t-\hat{\boldsymbol{\lambda}})A(\mathbf{x}^t-\mathbf{x}^{t+1})+||\mathbf{y}^{t+1}-\tilde{\mathbf{y}}||_2^2/\sigma\\
&\frac{1}{\alpha}(\hat{\boldsymbol{\lambda}}-\tilde{\boldsymbol{\lambda}})^*(\boldsymbol{\lambda}^t-\hat{\boldsymbol{\lambda}})+\frac{\alpha}{\tau}(\mathbf{x}^{t+1}-\tilde{\mathbf{x}})^*(\mathbf{x}^t-\mathbf{x}^{t+1})\\
&\geq(\boldsymbol{\lambda}^t-\hat{\boldsymbol{\lambda}})A(\mathbf{x}^t-\mathbf{x}^{t+1})
\end{aligned}\tag{7.23}$$

For convenience, if we define I_n the identity matrix of order n, and

$$G_0=\begin{pmatrix} I_n & \\ & \eta I_m\end{pmatrix},G_1=\begin{pmatrix}\frac{\alpha}{\tau}I_n & \\ & \frac{1}{\alpha}I_m\end{pmatrix},G=\begin{pmatrix}\frac{\alpha}{\tau}I_n & \\ & \frac{1}{\alpha\eta}I_m\end{pmatrix},$$

$$\hat{\mathbf{x}}=\mathbf{x}^{t+1},\quad \mathbf{u}=\begin{pmatrix}\mathbf{x}\\ \boldsymbol{\lambda}\end{pmatrix},\quad ||u||_G^2:=u^*Gu,$$

then (7.23) can be rewritten as

$$(\mathbf{u}^t-\tilde{\mathbf{u}})^*G_1(\mathbf{u}^t-\hat{\mathbf{u}})\geq||\mathbf{u}^t-\hat{\mathbf{u}}||_{G_1}^2+(\boldsymbol{\lambda}^t-\hat{\boldsymbol{\lambda}})^*A(\mathbf{x}^t-\hat{\mathbf{x}}).\tag{7.24}$$

The iteration of $\mathbf{u}$ in (7.18) is now expressed as $\mathbf{u}^{t+1}=\mathbf{u}^t-G_0(\mathbf{u}^t-\hat{\mathbf{u}})$, and it can be shown that

$$\begin{aligned}
&||\mathbf{u}^t-\tilde{\mathbf{u}}||_G^2-||\mathbf{u}^{t+1}-\tilde{\mathbf{u}}||_G^2\\
=\;&2(\mathbf{u}^t-\tilde{\mathbf{u}})^*G_1(\mathbf{u}^t-\hat{\mathbf{u}})-||G_0(\mathbf{u}^t-\hat{\mathbf{u}})||_G^2\\
\text{from (7.24)}\;\geq\;&2||\mathbf{u}^t-\hat{\mathbf{u}}||_{G_1}^2+2(\boldsymbol{\lambda}^t-\hat{\boldsymbol{\lambda}})^*A(\mathbf{x}^t-\hat{\mathbf{x}})-||\mathbf{u}^t-\hat{\mathbf{u}}||_{G_0GG_0}^2\\
=\;&\frac{\alpha}{\tau}||\mathbf{x}^t-\hat{\mathbf{x}}||_2^2+\frac{2-\eta}{\alpha}||\boldsymbol{\lambda}^t-\hat{\boldsymbol{\lambda}}||_2^2+2(\boldsymbol{\lambda}^t-\hat{\boldsymbol{\lambda}})^*A(\mathbf{x}^t-\hat{\mathbf{x}})
\end{aligned}$$

$$\begin{aligned} &\geq \frac{\alpha}{\tau}||\mathbf{x}^t - \hat{\mathbf{x}}||_2^2 + \frac{2-\eta}{\alpha}||\boldsymbol{\lambda}^t - \hat{\boldsymbol{\lambda}}||_2^2 - \zeta||\boldsymbol{\lambda}^t - \hat{\boldsymbol{\lambda}}||_2^2 - \frac{1}{\zeta}||A(\mathbf{x}^t - \hat{\mathbf{x}})||_2^2 \\ &\geq (\frac{\alpha}{\tau} - \frac{\lambda_{\max}}{\zeta})||\mathbf{x}^t - \hat{\mathbf{x}}||_2^2 + (\frac{2-\eta}{\alpha} - \zeta)||\boldsymbol{\lambda}^t - \hat{\boldsymbol{\lambda}}||_2^2 \end{aligned}$$

where ζ is a positive parameter and $\lambda_{\max}$ denotes the maxmimum eigenvalue of A^*A.

To keep on the convergence analysis, it must be satisfied that:

$$\begin{cases} \frac{\alpha}{\tau} - \frac{\lambda_{\max}}{\zeta} & > 0 \\ \frac{2-\eta}{\alpha} - \zeta & > 0 \end{cases}$$

which leads to

$$\tau\lambda_{\max} + \eta < 2. \tag{7.25}$$

With the condition in (7.25) and the assumption $\kappa = 1 - \tau\lambda_{\max}/(2-\eta) > 0$, $\zeta = (2-\eta)/(\alpha + \alpha\kappa) > 0$, we have

$$\begin{aligned} ||\mathbf{u}^t - \tilde{\mathbf{u}}||_G^2 - ||\mathbf{u}^{t+1} - \tilde{\mathbf{u}}||_G^2 &\geq \frac{\alpha\kappa^2}{\tau}||\mathbf{x}^t - \hat{\mathbf{x}}||_2^2 + \frac{2-\eta}{\alpha}\frac{\kappa}{1+\kappa}||\boldsymbol{\lambda}^t - \hat{\boldsymbol{\lambda}}||_2^2 \\ &\geq \epsilon||\mathbf{u}^t - \mathbf{u}^{t+1}||_G^2 \end{aligned} \tag{7.26}$$

where $\epsilon = \min(\kappa^2, \frac{\kappa(2-\eta)}{\eta(1+\kappa)}) > 0$. We can conclude from (7.26) that:

1. $||\mathbf{u}^t - \mathbf{u}^{t+1}||_G^2$ converges to 0, and thus $\lim_{t\to\infty} A\mathbf{x}^t + \mathbf{y}^t = \mathbf{b}$;
2. $\{\mathbf{u}^t\}$ lies in a compact region;
3. $||\mathbf{u}^t - \tilde{\mathbf{u}}||_G^2$ is monotonically non-increasing and therefore converges.

7.4 Simulation

7.4.1 Sinogram inpainting

The sinograms of three images in Figure 5.4 are under-sampled randomly and then interpolated by different methods. Comparisons of the simulation results of sinogram inpainting in terms of the normalized mean square error (NMSE) were made between other interplolation methods and our CS based model, and the results are shown in Figure 7.2.

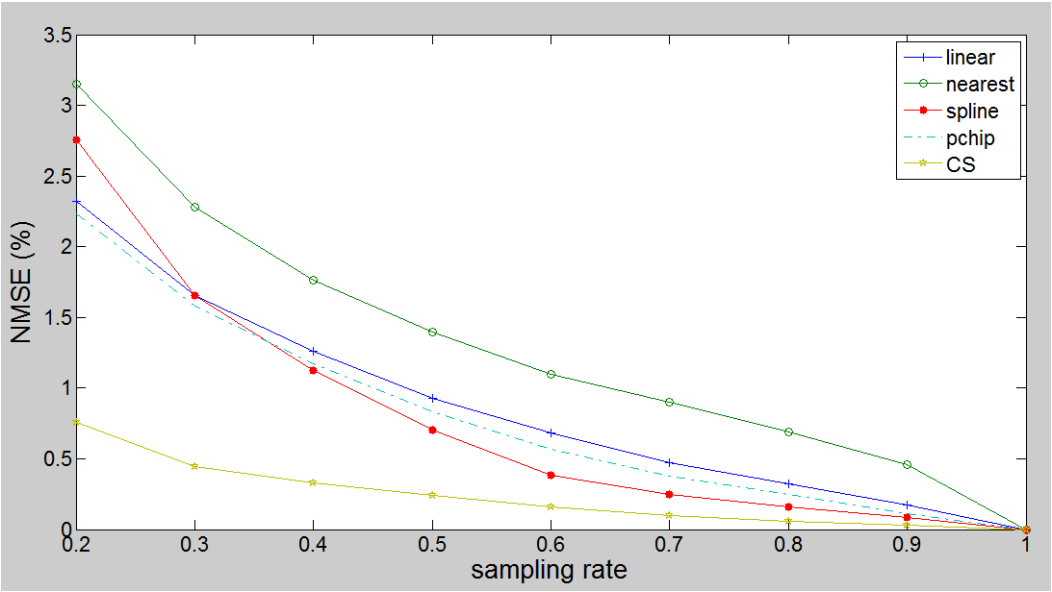
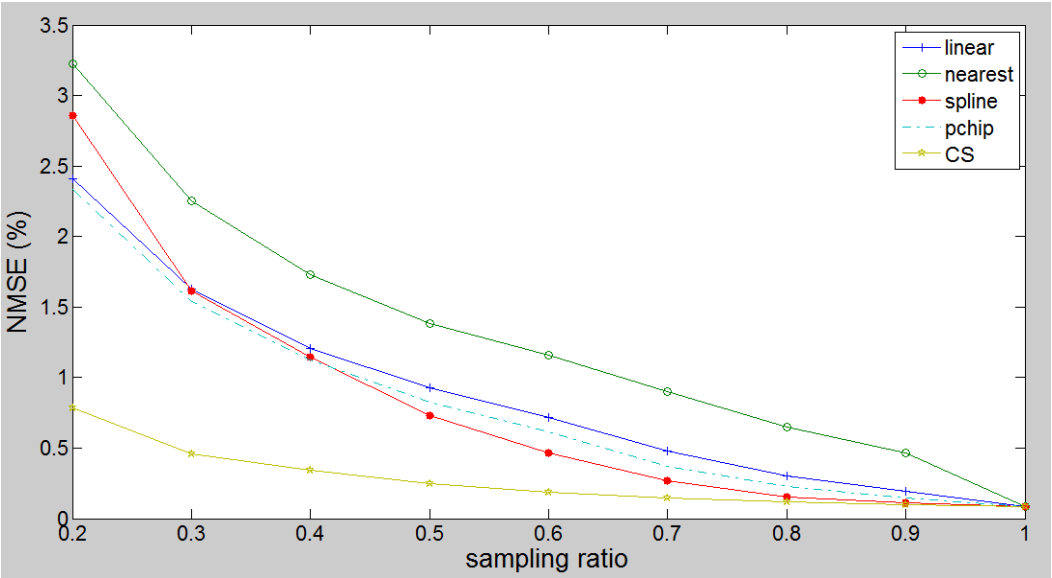


Figure 7.2: Comparisons of sinogram inpainting results

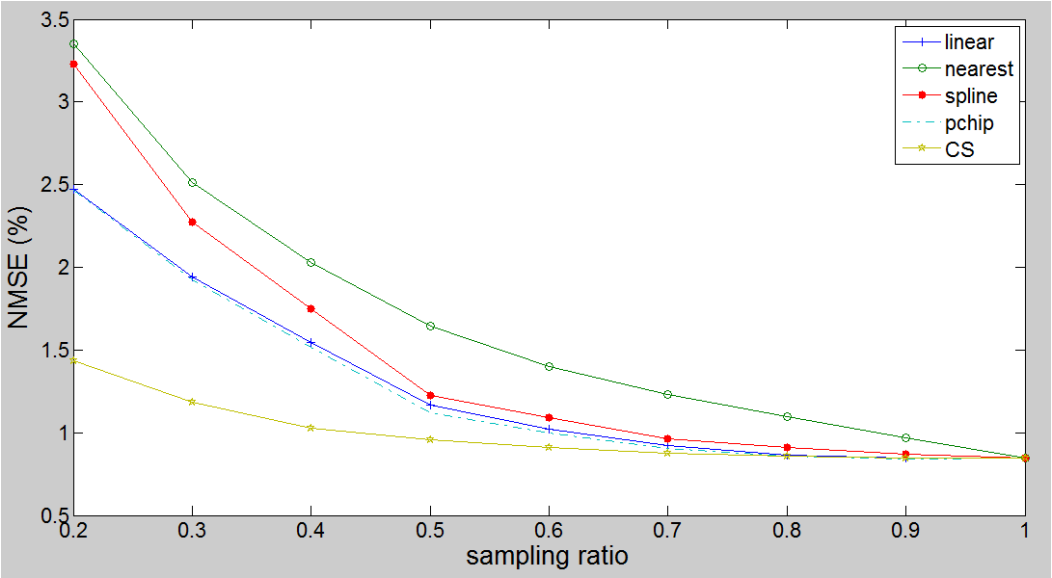
Minimal levels of noise were introduced to assess the stability of the algorithm. Different levels of Gaussian noise were added where the variance was proportional to the line integral values (a reasonable approximation for CT). The noise of variance (0.1%, 1%) is adopted here to compare the inpainting robustness in the presence of noise. The results are shown in Figure 7.3. From Figure 7.2 and Figure 7.3 we can tell that the proposed model is the most robust and outperforms other inpainting methods.

7.4.2 Object reconstruction

This section presents the computation results of the proposed CS reconstruction algorithm on simulated fan-beam sinogram data sets. Three simulation examples were computed. The first example is the Shepp-Logan phantom data set, and the second and third examples are CT image data sets. The proposed $w\ell_1$ regularized optimization algorithm and the existing CSTV algorithm [14] were applied respectively to the three phantom and image data sets and their reconstruction results were compared visually and quantitatively, in terms of the least squares error. The experimental setting and fan-beam scan parameters were as follows: the object was of size 256×256, $D = 256, \gamma = [-45 : 0.25 : 45], \beta$ were 180 random angles in the range $[0 : 1 : 359]$. For the $w\ell_1$ algorithm, the fan projections were converted at $\theta = [0 : 0.5 : 179.5]$. The value of the regularization parameter



(a) *noise of variance 0.1%*



(b) *noise of variance 1%*

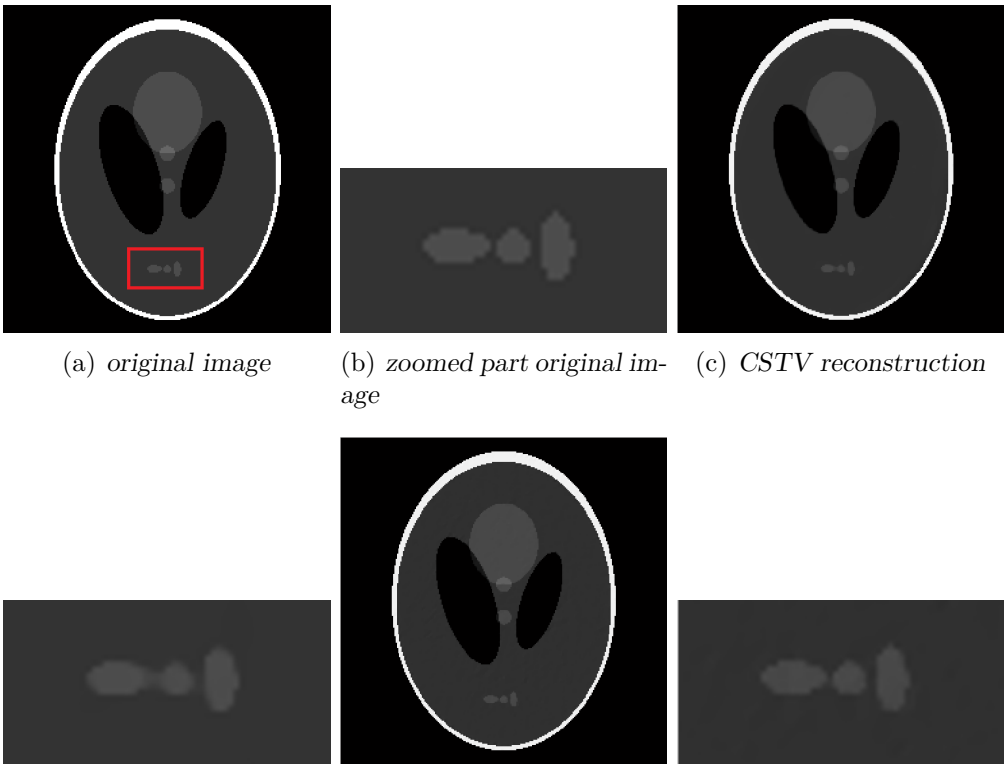
Figure 7.3: Comparisons of sinogram inpainting results in the presence of noise

α for different data sets was empirically selected. The reconstruction images of the three examples are shown in Figs. 7.4-7.6, respectively. To compare some details of the reconstructed images, some parts are zoomed and displayed.

NMSE and SNR defined in Chapter 5 and 6 were employed to provide quantitative evaluation and comparison of the reconstructed images. Table 7.1 provides a list of the NMSE and SNR values of the images and those of their selected zoomed

parts reconstructed by the proposed $w\ell_1$ regularized optimization algorithm and the existing CSTV algorithm.

Both the visual image display and quantitative values demonstrate that the proposed $w\ell_1$ regularized optimization algorithm provides better image reconstruction performance than the CSTV algorithm.



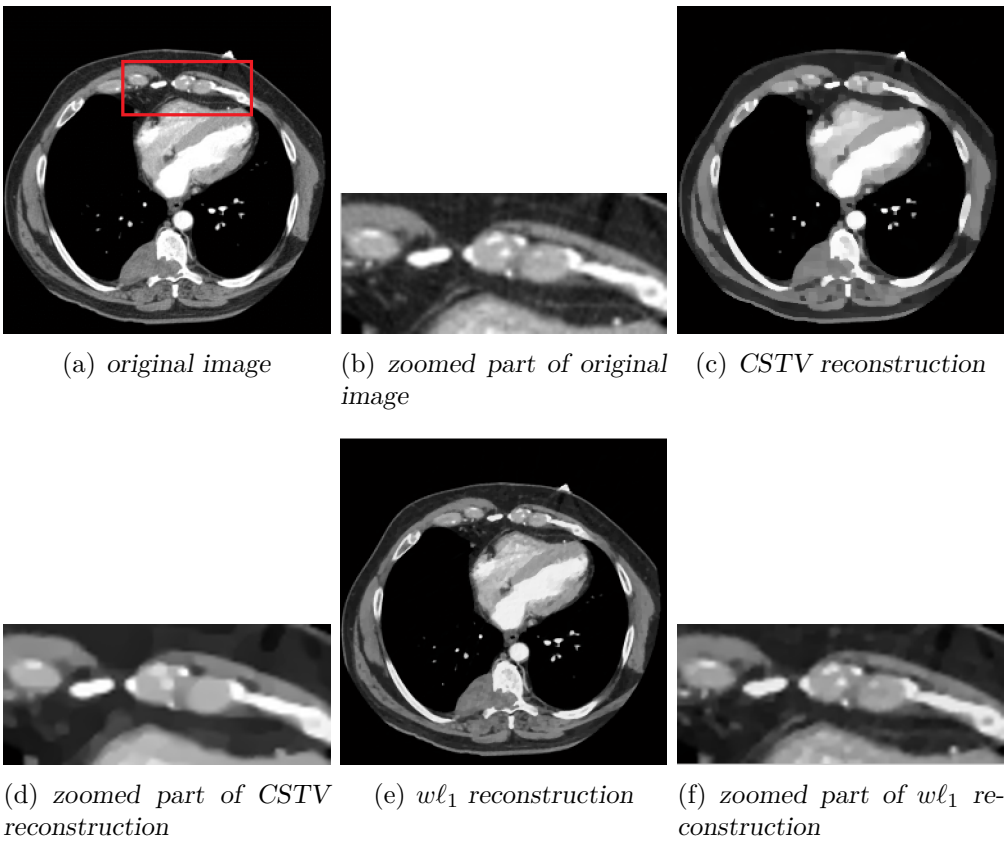
(a) *original image* (b) *zoomed part original image* (c) *CSTV reconstruction*

(d) *zoomed part of CSTV reconstruction* (e) *$w\ell_1$ reconstruction* (f) *zoomed part of $w\ell_1$ reconstruction*

Figure 7.4: Reconstruction of Shepp-Logan phantom

Table 7.1: Comparison of reconstructions in terms of NMSE and SNR

		Shepp-Logan		1st CT		2nd CT	
		NMSE	SNR	NMSE	SNR	NMSE	SNR
whole	by CSTV	5.82%	33.81	9.95%	28.14	8.48%	24.99
image	by $w\ell_1$	5.58%	34.03	6.89%	28.97	6.62%	26.77
zoomed	by CSTV	7.54%	25.21	10.06%	20.12	13.60%	17.73
part	by $w\ell_1$	6.28%	26.40	7.38%	21.59	7.26%	19.18



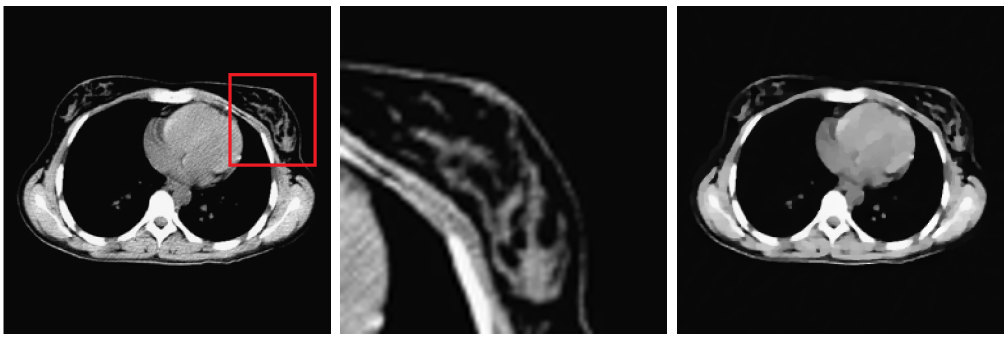
(a) *original image* (b) *zoomed part of original image* (c) *CSTV reconstruction*

(d) *zoomed part of CSTV reconstruction* (e) *$w\ell_1$ reconstruction* (f) *zoomed part of $w\ell_1$ reconstruction*

Figure 7.5: Reconstruction of CT image 1

7.5 Conclusion

In this chapter, we have analyzed the indirect application of CS in CT reconstruction. A CS approach to the fan-beam CT image reconstruction has been proposed for the purpose of reducing the X-ray radiation dosage in CT examinations without compromising image quality. The proposed approach is novel in the sense that it overcomes the difficulties of the existing CS approaches to CT image reconstruction which have not been able to establish the incoherence and sparsity conditions required by the CS theory for signal reconstruction. Instead of applying CS to the reconstruction of the object directly from the sinogram data set, the proposed CS approach indirectly reconstructs the sparse or compressible DFT of the sinograms via the optimization algorithm of yall1. The correspondingly formulated sensing scheme results in a randomly sampled partial DFT matrix which meets the theoretical RIP condition for CS reconstruction. The proposed weighted ℓ_1 regularized optimization algorithm takes into account the low pass property of the reconstruction signal and its reconstruction results outperform



(a) *original image* (b) *zoomed part of original image* (c) *CSTV reconstruction*

(d) *zoomed part of CSTV reconstruction* (e) $w\ell_1$ *reconstruction* (f) *zoomed part of* $w\ell_1$ *reconstruction*

Figure 7.6: Reconstruction of CT image 2

the existing CSTV reconstruction method. The proposed CS approach has so far been applied only to the 2-D fan-beam CT data sets and reconstruction in the noise free case. Its concept and formulation can be further extended to 3-D cone-beam CT data sets and image reconstruction with measurement noise.

Chapter 8

Conclusion and Future Work

8.1 Conclusion

X-ray computed tomography (CT) is a popular medical imaging modality but is faced with the dilemma of keeping reconstruction quality high and radiation low. The emerging compressed sensing (CS) technique, on the other hand, can recover signals from far fewer samples than traditionally required by the Shannon-Nyquist sampling theorem. The application of CS in CT is able to reduce the X-ray radiation dose to patients while preserving imaging quality, and has therefore been a topic of interest in the medical imaging community. However, there is a fundamental lack of understanding if the system matrix of CT can satisfy the restricted isometry property (RIP), an essential condition for successful CS reconstruction. In this thesis, we carry out the work on the application of CS in CT, aiming to bridge the existing gap between the theoretical results of CS and practical CT settings.

After reviewing the fundamentals of CS and CT in detail, we proposed the Fourier slice theorem (FST)-based CT model in Chapter 4. The object is to connect CT with Fourier transform, of which the RIP is well defined. With the preliminary knowledge of FST and the relation between different Fourier transforms, the projection model of parallel-beam CT was established and the error analysis indicated that the sinogram error becomes negligible when the padding size is sufficiently large. We continued to apply this model to fan-beam CT and depicted fully what the model is like for full-scanning, sparse-view and sparse-detector fan-

beam CT, respectively. The model building in this chapter laid the foundations for the subsequent analysis.

Chapter 5 presented the core investigation of the incoherence property of the CT data scanning schemes. We have analyzed which type of under-sampling is favourable to CS reconstruction. The work was conducted based on point spread function (PSF) and FST, respectively. The Frobenius norm of the difference matrix between PSF and the identity matrix was proposed as an evaluation index, and was shown experimentally to be more sensitive than another index, the sidelobe-to-peak ratio. Based on FST, the system matrix was decomposed to the Fourier matrix and our problem was converted to the study of sampling on the Fourier matrix, which involved the equivalence of the synthesis and analysis model, the viability of CS with coherent and redundant basis, and finally the RIP of distinct block diagonal matrices. The conclusion was drawn that golden-angle and random-detector scanning is the most favourable for post-CS-based reconstruction.

We further analysed the application of CS in CT by proposing two novel reconstruction models for the reduction of X-ray radiation dose in CT examinations without compromising image quality by CS. In Chapter 6, the CT projections were first converted to Mojette projections, an exact discretization of the Radon transform. On each view angle, the projections were summed up with the application of certain principles. 1-D Fourier transform was then applied to the merged Mojette projections, followed by an exact mapping of the gained Fourier coefficients to the 2-D Fourier domain of the scanning object. Finally the scanning object was reconstructed from the partial Fourier coefficients by CS with mixed constraints of TV and wavelet minimization.

In Chapter 7, the smoothness of the sinogram data set and the sparsity of its frequency transformation were exploited which, in conjunction with a randomly projected X-ray radiation scheme, can result in a randomly under-sampled partial Fourier matrix as the sensing matrix for CS reconstruction. The formulation of this CT data acquisition and reconstruction scheme satisfies the incoherence and sparsity properties required by CS theory. Based on this scheme, a weighted ℓ_1 regularized optimization algorithm is proposed for computing the CS image reconstruction. The reconstruction performance and advantages over other known CT reconstruction methods are demonstrated by simulated phantom and CT

images.

8.2 Future Work

We list some possible directions for future research.

- Extension of CS to statistical CT reconstruction in theory and implementation

  In Chapter 4 and 5, we have investigated the direct application of CS in CT reconstruction from theoretical and implementation perspectives. Although the algorithms exploited here is closely related to the ART-based problem, it is stated in Section 2.4.2 that the statistical reconstruction technique can potentially obtain better results as it considers the noise during the scanning process. This inspired the idea of extending CS to statistical CT reconstruction. While some attempts have been made to incorporate CS within the framework of statistical reconstructions [33, 87–89], no-one has tackled the RIP condition of the statistical formulation to the best of our knowledge. Compared to the ART-based model, the statistical model is more complicated, which makes the application of CS more challenging and interesting for future work.

- Application of CS to 3-D cone-beam CT

  The current work has been mainly done in 2-D and a straightforward idea for future work would be to extend the results to 3-D cone-beam CT (CBCT), which is becoming increasingly important in treatment planning and diagnosis. We have done some initial work on CBCT. For example, the formula and implementation of FDK and the Katsevich algorithm have been studied, and some simulations have been conducted to validate the theory. After building the system matrix, we have also tried to do the reconstruction based on CS in two ways. One is to reconstruct layer by layer, and the other is to deal with the voxel data. The first method is essentially 2-D and easy to implement, but at the same time quite slow as each layer is reconstructed separately, and some measurements are neglected and wasted because they involve different layers. In contrast, the other method is more promising, which applies CS directly to the 3-D system by minimizing the TV norm

of the 3-D object. The specific implementation can be found in [67]. The difficulty with the latter method is that the system matrix is massive for storage and calculation at each iteration. In addition to the implementation of CS, we are also interested in the theoretical analysis. 3-D CT scanning has a higher degree of freedom as there are more parameters, which leaves us abundant choices of different sampling schemes. Similar to the work that we have done for 2-D CT scanning, it would be beneficial to find out which schemes are favourable for post-CS-based reconstruction. In other words, it is important to know how the radiation can be reduced while affecting quality the least. Hence, the future work is to continue with the study of CS in 3-D CBCT.

- CT reconstruction on GPU

  With progress in imaging systems and algorithms, the computational complexity of image reconstruction has increased dramatically, which leads to increased processing time for current reconstruction techniques. However, fast image reconstruction is often required to deliver a prompt diagnosis and even more critical in real-time imaging applications, such as intraoperative cone-beam CT. The graphics processing unit (GPU) has emerged as a competitive platform for computing massively parallel problems [120]. To exploit the capabilities of GPU, CT reconstruction problems can be formulated as data-parallel tasks for efficient implementation. Since the analytical reconstruction is fully parallel, the accelerated versions of FBP for 2-D images [53] and FDK for 3-D images [111, 114] have been implemented successfully. In contrast, iterative reconstruction is fundamentally sequential, therefore computationally challenging for GPU acceleration. For example, the algebraic reconstruction technique (ART) cannot be implemented efficiently on GPU because each iteration only processes a single projection line. Some efforts have been made towards adapting iterative algorithms on GPU, such as simultaneous ART [104], expectation maximization (EM) [31, 156], ordered-subsets EM [31, 157], and TV minimization reconstruction [70]. Inspired by these precursory studies, the future research direction can be to implement the CS-based algorithms developed in this thesis on GPU, aiming at highly efficient and accurate reconstructions. The sparse property of the CT system matrix can be favourable for implementation, because the efficient matrix-vector multiplication is performed implicitly with a procedural algorithm rather than explicitly with

the matrix multiplication formula.

References

[1] NCRP Report No.160: ionizing radiation exposure of the population of the United States. `http://www.ncrponline.org/Publications/Press_Releases/160press.html`, 2009.

[2] S. Abbas, T. Lee, S. Shin, R. Lee, and S. Cho. Effects of sparse sampling schemes on image quality in low-dose CT. *Medical Physics*, 40:111915, 2013.

[3] S. Abbas, J. Min, and S. Cho. Super-sparsely view-sampled cone-beam CT by incorporating prior data. *J. X-Ray Sci. Technol.*, 21:71–83, 2013.

[4] F. Affentranger and R. Schneider. Random projections of regular simplices. *Discrete Comput. Geom.*, 7:219–226, 1992.

[5] A. H. Andersen and A. C. Kak. Simultaneous algebraic reconstruction technique (SART): a superior implementation of the ART algorithm. *Ultrason. Imaging*, 6:81–94, 1984.

[6] A. Averbuch, D. Donoho, R. Coifman, M. Israeli, and J. Walden. *Fast slant stack: A notion of Radon transform for data on a cartesian grid which is rapidly computable, algebraically exact, geometrically faithful and invertible.* Department of Statistics, Stanford University, 2001.

[7] C. Badea and R. Gordon. Experiments with the nonlinear and chaotic behaviour of the multiplicative algebraic reconstruction technique (MART) algorithm for computed tomography. *Phys Med Biol*, 49:1455–1474, 2004.

[8] J. Barzilai and J. M. Borwein. Two-point step size gradient methods. *IMA J. Numer. Anal.*, 8:141–148, 1988.

[9] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imag. Sci.*, 2:183–202, 2008.

[10] M. Beister, D. Kolditz, and W. A. Kalender. Iterative reconstruction methods in X-ray CT. *Physica Med.*, 28:94–108, 2012.

[11] G. A. Berrington, M. Mahesh, K. P. Kim, M. Bhargavan, R. Lewis, F. Mettler, and C. Land. Projected cancer risks from computed tomographic scans

performed in the United States in 2007. *Arch. Intern. Med.*, 169:2071–2077, 2009.

[12] G. Beylkin. Discrete radon transform. *IEEE Trans. on Acoustics, Speech, Signal Processing*, 35:162–172, 1987.

[13] J. Bian, J. H. Siewerdsen, X. Han, E. Y. Sidky, J. L. Prince, C. A. Pelizzari, and X. Pan. Evaluation of sparse-view reconstruction from flat-panel-detector cone-beam CT. *Phys. Med. Biol.*, 55:6575–6599, 2010.

[14] J. Bian, J. Wang, X. Han, E. Y. Sidky, L. Shao, and X. Pan. Optimization-based image reconstruction from sparse-view data in offset-detector CBCT. *Physics in Medicine and Biology*, 58:205–230, 2013.

[15] T. Blumensath and M. Davies. Iterative hard thresholding for compressive sensing. *Appl. Comput. Harmon. Anal.*, 27:265–274, 2009.

[16] F. E. Boas and D. Fleischmann. CT artifacts: Causes and reduction techniques. *Imaging Med.*, 4:229–240, 2012.

[17] M. L. Brady. A fast discrete approximation algorithm for the radon transform. *SIAM Journal on Computing*, 27:107–119, 1998.

[18] D. J. Brenner and E. J. Hall. Computed tomography-an increasing source of radiation exposure. *N. Engl. J. Med.*, 357:2277–2284, 2007.

[19] K. Bryan and T. Leise. Making do with less: An introduction to compressed sensing. *SIAM Review*, 55:547–566, 2013.

[20] E. J. Candès, Y. C. Eldarb, D. Needell, and P. Randall. Compressed sensing with coherent and redundant dictionaries. *Appl. Comput. Harmon. Anal.*, 31:59–73, 2011.

[21] E. J. Candès and J. Romberg. Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE Transactions on Information Theory*, 52:5406–5425, 2006.

[22] E. J. Candès and J. Romberg. Quantitative robust uncertainty principles and optimally sparse decompositions. *Foundations of Computational Mathematics*, 6:227–254, 2006.

[23] E. J. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, 52:489–509, 2006.

[24] E. J. Candès and M. B. Wakin. An introduction to compressive sampling. *IEEE Signal Processing Magazine*, 25:21–30, 2008.

[25] E. J. Candès, M. B. Wakin, and S. Boyd. Enhancing sparsity by reweighted l1 minimization. *Journal of Fourier Analysis and Applications*, 14:877–905, 2008.

[26] R. W. Chan, E. A. Ramsay, E. Y. Cheung, and D. B. Plewes. The influence of radial undersampling schemes on compressed sensing reconstruction in breast MRI. *Magnetic Resonance in Medicine*, 67:363–377, 2012.

[27] S. Chandra, I. Svalbe, J. Guedon, A. Kingston, and N. Normand. Recovering missing slices of the discrete fourier transform using ghosts. *IEEE Transactions on Image Processing*, 21:4431 – 4441, 2012.

[28] S. S. Chandra, N. Normand, A. Kingston, J. Guédon, and I. Svalbe. Fast Mojette transform for discrete tomography. Preprint, `http://arxiv.org/abs/1006.1965v1`, 2010.

[29] G. H. Chen, J. Tang, and S. H. Leng. Prior image constrained compressed sensing (PICCS): A method to accurately reconstruct dynamic CT images from highly undersampled projection data sets. *Medical Physics*, 35:660–673, 2008.

[30] S. S. Chen, D. L. Donoho, and M. A. Saunders. Atomic decomposition by basis pursuit. *SIAM J. Sci. Comput.*, 20:33–61, 1999.

[31] K. Chidlow and T. Moller. Rapid emission tomography reconstruction. In *Proceedings of the Eurographics*, pages 15–26, New York, USA, 2003.

[32] S. Cho, T. Lee, J. Min, and H. Chung. Feasibility study on many-view under sampling (MVUS) technique for low-dose computed tomography. *Opt. Eng.*, 51:080501, 2012.

[33] K. Choi, J. Wang, L. Zhu, T. S. Suh, S. Boyd, and L. Xing. Compressed sensing based cone-beam computed tomography reconstruction with a first-order method. *Med. Phys.*, 37:5113–5125, 2010.

[34] J. Christner, V. Zavaletta, C. Eusemann, A. I. Walz-Flannigan, and C. H. McCollough. Dose reduction in spiral CT using dynamically adjustable z-axis x-ray beam collimation. *AJR*, 194:W49–W55, 2010.

[35] W. Dai and O. Milenkovic. Subspace pursuit for compressive sensing signal reconstruction. *IEEE Trans. Inform. Theory*, 55:2230–2249, 2009.

[36] M. Davenport and M. Wakin. Analysis of orthogonal matching pursuit using the restricted isometry property. *IEEE Trans. Inform. Theory*, 56:4395–4401, 2010.

[37] M. A. Davenport, M. F. Duarte, Y. C. Eldar, and G. Kutyniok. Introduction to compressed sensing. In Y. C. Eldar and G. Kutyniok, editors, *Compressed Sensing: Theory and Applications*, pages 1–64. Cambridge University Press, 2012.

[38] F. Dennerlein, F. Noo, J. Hornegger, and G. Lauritsch. Fan-beam filtered-backprojection reconstruction without backprojection weight. *Physics in Medicine and Biology*, 52:3227–3240, 2007.

[39] D. Donoho and Y. Tsaig. Fast solution of ℓ_1 norm minimization problems when the solution may be sparse. *IEEE Trans. Inform. Theory*, 54:4789–4812, 2008.

[40] D. L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52:1289–1306, 2006.

[41] D. L. Donoho. High-dimensional centrally symmetric polytopes with neighborliness proportional to dimension. *Discrete Comput. Geom.*, 35:617–652, 2006.

[42] D. L. Donoho and J. Tanner. Neighborliness of randomly projected simplices in high dimensions. *Proc. Natl. Acad. Sci.*, 102:9452–9457, 2005.

[43] D. L. Donoho and J. Tanner. Counting faces of randomly-projected polytopes when the projection radically lowers dimension. *J. Amer. Math. Soc.*, 22:1–53, 2009.

[44] D. L. Donoho, Y. Tsaig, I. Drori, and J. L. Starck. Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit. Technical report, Dept. Statistics, Stanford Univ., Stanford, CA, 2006.

[45] D. L. Donoho, Y. Tsaig, I. Drori, and J. L. Starck. Sparse solution of underdetermined systems of linear equations by stagewise orthogonal matching pursuit. *IEEE Trans. Inform. Theory*, 58:1094–1121, 2012.

[46] M. Duarte, M. Davenport, D. Takhar, J. Laska, S. Ting, K. Kelly, and R. Baraniuk. Single-pixel imaging via compressive sampling. *IEEE Signal Processing Magazine*, 25:83–91, 2008.

[47] M. Elad, P. Milanfar, and R. Rubinstein. Analysis versus synthesis in signal priors. *Inverse Problems*, 23:947–968, 2007.

[48] H. Fayad, J. P. Guédon, I. Svalbe, Y. Bizais, and N. Normand. Applying Mojette discrete radon transforms to classical tomographic data. In *Proc. of SPIE*, volume 6913, 2008.

[49] L. A. Feldkamp, L. C. Davis, and J. W. Kress. Practical cone beam algorithm. *J. Opt. Soc. Am. A*, 1:612–619, 1984.

[50] M. Figueiredo, R. Nowak, and S.Wright. Gradient projections for sparse reconstruction: Application to compressed sensing and other inverse problems. *IEEE J. Select. Top. Signal Processing*, 1:586–597, 2007.

[51] M. Fornasier and H. Rauhut. Compressive sensing. In O. Scherzer, editor, *Handbook of Mathematical Methods in Imaging*. Springer, New York, 2011.

[52] J. Friedman, T. Hastie, and R. Tibshirani. Regularization paths for generalized linear models via coordinate descent. *J. Stats. Software*, 33:1–22, 2010.

[53] M. Garland, S. L. Grand, J. Nickolls, J. Anderson, J. Hardwick, S. Morton, E. Phillips, Y. Zhang, and V. Volkov. Parallel computing experiences with CUDA. *IEEE Micro*, 28:13–27, 2008.

[54] R. Glowinski. *Numerical methods for nonlinear variational problems*. Springer Verlag, 1984.

[55] R. Glowinski and P. L. Tallec. *Augmented Lagrangian and operator splitting methods in nonlinear mechanics*. 1989.

[56] T. Goldstein and S. Osher. The split Bregman method for l1-regularized problems. *SIAM J. Imaging Sci.*, 2:323–343, 2009.

[57] R. Gordon, R. Bender, and G. T. Herman. Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and x-ray photography. *J Theor Biol*, 29:471–482, 1970.

[58] W. Götz and H. Druckmüller. A fast digital radon transform-an efficient means for evaluating the hough transform. *Pattern Recognition*, 29:711–718, 1996.

[59] J. P. Guédon, D. Barba, and N. Burger. Psychovisual image coding via an exact discrete radon transform. In *Proc. of the SPIE - The International Society for Optical Engineering*, volume 2501, pages 562–572, 1995.

[60] J. Hale. *The fundamentals of radiological science*. Springfield, IL: Charles C. Thomas, 1974.

[61] X. Han, J. Bian, and D. R. Eaker. Algorithm-enabled low-dose micro-CT imaging. *IEEE Trans. Medical Imaging*, 30:606–620, 2011.

[62] X. Han, J. Bian, E. L. Ritman, E. Y. Sidky, and X. Pan. Optimization based reconstruction of sparse images from few-view projections. *Phys. Med. Biol.*, 57:5245–5273, 2012.

[63] B. Hayes. The best bits. *Amer. Scientist*, 97:276–280, 2009.

[64] G. T. Herman. *Image Reconstruction from Projections, the Fundamentals of Computerized Tomography*. New York: Academic Press, 1980.

[65] M. Herman and T. Strohmer. High-resolution radar via compressed sensing. *IEEE Trans. Signal Process.*, 57:2275–2284, 2009.

[66] F. Holmquist and U. Nyman. Eighty-peak kilovoltage 16-channel multidetector computed tomography and reduced contrast-medium doses tailored to body weight to diagnose pulmonary embolism in azotaemic patients. *Eur. Radiol.*, 16:1165–1176, 2006.

[67] T. L. Jensen, J. H. Jørgensen, P. C. Hansen, and S. H. Jensen. Implementation of an optimal first-order method for strongly convex total variation regularization. *BIT Numerical Mathematics*, 52:329–356, 2011.

[68] X. Jia, B. Dong, Y. Lou, and S. B. Jiang. GPU-based iterative cone beam CT reconstruction using tight frame regularization. *Physics in Medicine and Biology*, 56:3787–3807, 2011.

[69] X. Jia, Y. Lou, J. Lewis, R. Li, X. Gu, C. Men, W. Y. Song, and S. B. Jiang. GPU-based fast low-dose cone beam CT reconstruction via total variation. *J. X-Ray Sci. Technol.*, 19:139–154, 2011.

[70] X. Jia, Y. Lou, R. Li, W. Y. Song, and S. B. Jianga. GPU-based fast cone beam CT reconstruction from undersampled and noisy projection data via total variation. *Medical Physics*, 37:1757–1760, 2010.

[71] M. Jiang and G. Wang. Convergence of the simultaneous algebraic reconstruction technique (SART). *IEEE Trans. Image Process*, 12:957–961, 2003.

[72] W. B. Johnson and J. Lindenstrauss. *Handbook of the Geometry of Banach Spaces Vol I.* North-Holland Publishing Co., 2001.

[73] J. H. Jørgensen, E. Y. Sidky, and X. Pan. Quantifying admissible undersampling for sparsity-exploiting iterative image reconstruction in x-ray CT. *IEEE Transactions on Medical Imaging*, 32:460–473, 2013.

[74] J. S. Jørgensen, E. Y. Sidky, P. C. Hansen, and X. Pan. Quantitative study of undersampled recoverability for sparse images in computed tomography. *S I A M Journal on Scientific Computing*, 2012.

[75] S. Kaczmarz. Angenäherte Auflösung von Systemen linearer Gleichungen. *Bull. Acad. Pol. Sci. Lett.*, A35:355–357, 1937.

[76] Y. Kaganovsky, D. Li, A. Holmgren, H. Jeon, K. P. MacCabe, D. G. Politte, J. A. O'Sullivan, L. Carin, , and D. J. Brady. Compressed sampling strategies for tomography. *J. Opt. Soc. Am. A*, 31:1369–1394, 2014.

[77] A. C. Kak and M. Slaney. *Principles of computerized tomographic imaging.* Society for Industrial and Applied Mathematics, Philadelphia, 2001.

[78] W. A. Kalender. X-ray computed tomography. *Phys. Med. Biol.*, 51:R29–R43, 2006.

[79] A. Katsevich. Analysis of an exact inversion formula for spiral cone-beam CT. *Physics in Medicine and Biology*, 47:2583–2598, 2002.

[80] A. Katsevich. Microlocal analysis of an FBP algorithm for truncated spiral cone-beam data. *Fourier Anal. Appl.*, 8:407–425, 2002.

[81] A. Katsevich. Theoretically exact filtered backprojection-type inversion algorithm for spiral CT. *SIAM Journal of Applied Mathematics*, 62:2012–2026, 2002.

[82] A. Katsevich. An improved exact filtered backprojection algorithm for spiral computed tomography. *Advances in Applied Mathematics*, 32:681–697, 2004.

[83] M. Katz. *Questions of Uniqueness and Resolution in Reconstruction from projections*. Sprlnger Verlag, 1979.

[84] B. T. Kelley and V. K. Madisetti. The fast discrete radon transform–i: Theory. *IEEE Trans. on Image Processing*, 2:382–400, 1993.

[85] A. Kingston, H. Li, N. Normand, and I. Svalbe. Fourier inversion of the mojette transform. In *Discrete Geometry for Computer Imagery*, volume 8668, pages 275–284. Springer International Publishing, 2014.

[86] V. Kotelnikov. On the carrying capacity of the ether and wire in telecommunications. In *Izd. Red. Upr. Svyazi RKKA*, Moscow, Russia, 1933.

[87] F. Krzakala, M. Mézard, F. Sausset, Y. F. Sun, and L. Zdeborová. Statistical-physics-based reconstruction in compressed sensing. *Physical Review X*, 2:021005, 2012.

[88] P. T. Lauzier and G. H. Chen. Characterization of statistical prior image constrained compressed sensing. I. applications to time-resolved contrast-enhanced ct. *Medical Physics*, 39:5930–5948, 2012.

[89] P. T. Lauzier and G. H. Chen. Characterization of statistical prior image constrained compressed sensing (PICCS): II. application to dose reduction. *Medical Physics*, 40:021902, 2013.

[90] M. Ledoux and M. Talagrand. *Probability in Banach Spaces*. Springer, 1991.

[91] C. Li. An efficient algorithm for total variation regularization with applications to the single pixel camera and compressive sensing. Master's thesis, Rice University, 2009.

[92] C. Li. *Compressive Sensing for 3D Data Processing Tasks: Applications, Models and Algorithms*. PhD thesis, Rice University, 2011.

[93] I. Loris. On the performance of algorithms for the minimization of ℓ_1-penalized functions. *Inverse Problems*, 25:035008, 2009.

[94] M. Lustig, D. Donoho, and J. M. Pauly. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magnetic Resonance in Medicine*, 58:1182–1195, 2007.

[95] M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly. Compressed sensing MRI. *IEEE Signal Processing Magazine*, 25:72–82, 2008.

[96] D. MacKenzie. Compressed sensing makes every pixel count. *What's Happening in the Mathematical Sciences*, 7:114–127, 2009.

[97] A. S. Malik, T. S. Choi, and H. Nisar. *Depth Map and 3D Imaging Applications: Algorithms and Technologies*. IGI Global, Hershey PA, US, 2011.

[98] S. G. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Trans. Signal Process.*, 41:3397–3415, 1993.

[99] B. D. Man and S. Basu. Distance-driven projection and backprojection. In *IEEE Nuclear Science Symposium and Medical Imaging Conference*, pages 1477–1480, Norfolk, Virginia, 2002.

[100] B. D. Man and S. Basu. Distance-driven projection and backprojection: Extension to three dimensions and analysis. *Physics in Medicine and Biology*, 49:2463–2475, 2004.

[101] S. H. Manglos, G. M. Gagne, A. Krol, F. D. Thomas, and R. Narayanaswamy. Transmission maximum-likelihood reconstruction with ordered subsets for cone beam CT. *Phys Med Biol*, 40:1225–1241, 1995.

[102] J. D. Mathews, A. V. Forsythe, Z. Brady, M. W. Butler, S. K. Goergen, G. B. Byrnes, G. G. Giles, A. B. Wallace, P. R. Anderson, T. A. Guiver, P. McGale, T. M. Cain, J. G. Dowty, A. C. Bickerstaffe, and S. C. Darby. Cancer risk in 680 000 people exposed to computed tomography scans in childhood or adolescence: data linkage study of 11 million Australians. *BMJ*, 346:f2360, 2013.

[103] C. H. McCollough, A. N. Primak, N. Braun, J. Kofler, L. Yu, and J. Christner. Strategies for reducing radiation dose in CT. *Radiol. Clin. North Am.*, 47:27–40, 2009.

[104] K. Mueller and R. Yagel. Rapid 3-D cone-beam reconstruction with the simultaneous algebraic reconstruction technique (SART) using 2-D texture mapping hardware. *IEEE Trans. Med. Imaging*, 19:1227–1237, 2000.

[105] B. K. Natarajan. Sparse approximate solutions to linear systems. *SIAM J. Comput.*, 24:227–234, 1995.

[106] D. Needell and J. Tropp. CoSaMP: Iterative signal recovery from incomplete and inaccurate samples. *Appl. Comput. Harmon. Anal.*, 26:301–321, 2009.

[107] D. Needell and R. Vershynin. Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit. *Found. Comput. Math.*, 9:317–334, 2009.

[108] D. Needell and R. Vershynin. Signal recovery from incomplete and inaccurate measurements via regularized orthogonal matching pursuit. *IEEE J. Select. Top. Signal Processing*, 4:310–316, 2010.

[109] B. E. Nett, R. Brauweiler, W. Kalender, H. Rowley, and G. H. Chen. Perfusion measurements by micro-CT using prior image constrained compressed sensing (PICCS): initial phantom results. *Phys. Med. Biol.*, 55:2333–2350, 2010.

[110] T. Niu and L. Zhu. Accelerated barrier optimization compressed sensing (ABOCS) reconstruction for cone-beam CT: Phantom studies. *Medical Physics*, 39:4588–4598, 2012.

[111] P. B. Noel, A. M. Walczak, J. Xu, J. J. Corso, K. R. Hoffmann, and S. Schafer. GPU-based cone beam computed tomography. *Comput. Methods Programs Biomed.*, 98:271–277, 2010.

[112] F. Noo, J. Pack, and D. Heuscher. Exact helical reconstruction using native cone-beam geometries. *Physics in Medcine and Biology*, 48:3787–3818, 2003.

[113] H. Nyquist. Certain topics in telegraph transmission theory. *Trans. AIEE*, 47:617–644, 1928.

[114] Y. Okitsu, F. Ino, and K. Hagihara. High-performance cone beam reconstruction using CUDA compatible GPUs. *Parallel Comput.*, 36:129–141, 2010.

[115] S. Osher, Y. Mao, B. Dong, and W. Yin. Fast linearized Bregman iterations for compressive sensing and sparse denoising. *Comm. in Math. Sciences*, 8:93–111, 2010.

[116] J. C. Park, B. Song, J. S. Kim, S. H. Park, H. K. Kim, Z. Liu, T. S. Suh, and W. Y. Song. Fast compressed sensing-based CBCT reconstruction using Barzilai-Borwein formulation for application to on-line IGRT. *Medical Physics*, 39:1207– 1217, 2012.

[117] J. Y. Park, H. L. Yap, C. J. Rozell, and M. B. Wakin. Concentration of measure for block diagonal matrices with applications to compressive signal processing. *IEEE Trans. on Signal Processing*, 59:5859–5875, 2011.

[118] T. M. Peters. Algorithms for fast back- and re-projection in computed tomography. *IEEE Transactions on Nuclear Science*, 28:3641–3647, 1981.

[119] G. Pfander and H. Rauhut. Sparsity in time-frequency representations. *J. Fourier Anal. Appl.*, 16:233–260, 2010.

[120] G. Pratxa and L. Xing. GPU computing in medical physics: A review. *Medical Physics*, 38:2685–2697, 2011.

[121] F. V. R and S. H. C. Physicians' views of the relative importance of thirty medical innovations. *Health Affairs*, 20:30–42, 2001.

[122] J. Radon. On the determination of functions from their integrals along certain manifold. *Berichte Sächs. Akad. Wiss.*, 69:262–278, 1917.

[123] E. A. Rashed and H. Kudo. Statistical image reconstruction from limited projection data with intensity priors. *Phys Med Biol*, 57:2039–2061, 2012.

[124] H. Rauhut. Stability results for random sampling of sparse trigonometric polynomials. *IEEE Trans. Information Theory*, 54:5661–5670, 2008.

[125] H. Rauhut. Compressive sensing and structured random matrices. In M. Fornasier, editor, *Theoretical Foundations and Numerical Methods for Sparse Recovery*, pages 1–94. Comp. Appl. Math XX, 2011.

[126] B. Recur, P. Desbarats, and J. Domenger. Mojette reconstruction from noisy projections. In *2010 2nd International Conference on Image Processing Theory Tools and Applications (IPTA)*, pages 201–206, 2010.

[127] B. Recur, P. Desbarats, and J. P. Domenger. Radon and Mojette projections' equivalence for tomographic reconstruction using linear systems. In *WSCG*, pages 191–198, Plzen, Czech Republic, 2008.

[128] L. Ritschl, F. Bergner, C. Fleischmann, and M. Kachelriess. Improved total variation-based CT image reconstruction applied to clinical data. *Phys Med Biol*, 56:1545–1561, 2011.

[129] J. Romberg. Imaging via compressive sampling. *IEEE Signal Process. Magazine*, 25:14–20, 2008.

[130] M. Rostami, O. Michailovich, and Z. Wang. Image deblurring using derivative compressed sensing for optical imaging application. *IEEE Trans. Image Processing*, 21:3139–3149, 2012.

[131] M. Rudelson and R. Vershynin. On sparse reconstruction from fourier and gaussian measurements. *Comm. Pure Appl. Math.*, 61:1025–1045, 2008.

[132] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation noise removal algorithm. *Physica D*, 21:3139–3149, 1992.

[133] K. Sauer and C. Bouman. A local update strategy for iterative reconstruction from projections. *IEEE Trans Sig Proc*, 41:534–548, 1993.

[134] M. Servières, N. Normand, J. Guédon, and Y. Bizais. The Mojette transform: Discrete angles for tomography. *Electronic Notes in Discrete Mathematics*, 20:587– 606, 2005.

[135] C. Shannon. Communication in the presence of noise. *Proc. Institute of Radio Engineers*, 37:10–21, 1949.

[136] C. E. Shannon. Communication in the presence of noise. In *Proc. IRE,*, volume 37, pages 10–21, 1949.

[137] R. Siddon. Fast calculation of the exact radiological path length for a three-dimensional CT array. *Medical Physics*, 12:252–255, 1985.

[138] E. Y. Sidky, M. A. Anastasio, and X. Pan. Image reconstruction exploiting object sparsity in boundary-enhanced X-ray phase-contrast tomography. *Opt. Express*, 18:10404–10422, 2010.

[139] E. Y. Sidky, C. M. Kao, and X. C. Pan. Accurate image reconstruction from few-views and limited-angle data in divergent-beam CT. *Journal of X-Ray Science and Technology*, 15:119–139, 2006.

[140] E. Y. Sidky and X. Pan. Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization. *Physics in Medicine and Biology*, 53:4777–4807, 2008.

[141] J. Song, Q. H. Liu, G. A. Johnson, and C. T. Badea. Sparseness prior based iterative image reconstruction for retrospectively gated cardiac micro-CT. *Medical Physics*, 34:4476–4483, 2007.

[142] J. Sunnegardh and P. E. Danielsson. Regularized iterative weighted filtered backprojection for helical cone-beam CT. *Medical Physics*, 35:4173–4185, 2008.

[143] J. Tang, B. E. Nett, and G. H. Chen. Performance comparison between total variation (TV)-based compressed sensing and statistical iterative reconstruction algorithms. *Phys. Med. Biol.*, 54:5781–5804, 2009.

[144] G. Tauböck, F. Hlawatsch, D. Eiwen, and H. Rauhut. Compressive estimation of doubly selective channels in multicarrier systems: Leakage effects and sparsity-enhancing processing. *IEEE J. Sel. Top. Sig. Process.*, 4:255–271, 2010.

[145] J. B. Thibault, K. D. Sauer, C. A. Bouman, and J. A. Hsieh. Three-dimensional statistical approach to improved image quality for multislice helical CT. *Medical Physics*, 34:4526–4544, 2007.

[146] P. Toft. *The Radon Transform : Theory and Implementation.* PhD thesis, Technical University of Denmark, 1996.

[147] J. A. Tropp, J. N. Laska, M. F. Duarte, J. K. Romberg, and R. G. Baraniuk. Beyond nyquist: Efficient sampling of sparse bandlimited signals. *IEEE Trans. Inform. Theory*, 56:520–544, 2010.

[148] J. Trzasko and A. Manduca. Highly undersampled magnetic resonance image reconstruction via homotopic ℓ_0-minimization. *IEEE Trans. Med. Imaging*, 28:106–121, 2009.

[149] Y. Tsaig and D. L. Donoho. Extensions of compressed sensing. *Signal Processing*, 86:549 – 571, 2006.

[150] A. Waaijer, M. Prokop, B. K. Velthuis, C. J. Bakker, G. A. Kort, and M. S. Leeuwen. Circle of willis at CT angiography: dose reduction and image quality-reducing tube voltage and increasing tube current settings. *Radiology*, 242:832–839, 2007.

[151] G. Wang, T.-H. Lin, P.-C. Cheng, , and D. M. Shinozaki. A general cone-beam reconstruction algorithm. *IEEE Trans. on Medical Imaging*, 12:486–496, 1993.

[152] Z. Wen, W. Yin, D. Goldfarb, and Y. Zhang. A fast algorithm for sparse reconstruction based on shrinkage, subspace optimization and continuation. *SIAM J. on Sci. Comp.*, 32:1832–1857, 2010.

[153] E. Whittaker. On the functions which are represented by the expansions of the interpolation theory. In *Proc. Royal Soc.*, volume 35, pages 181–194, Edinburgh, 1915.

[154] D. Wu, L. Li, and L. Zhang. Feature constrained compressed sensing CT image reconstruction from incomplete data via robust principal component analysis of the database. *Phys. Med. Biol.*, 58:4047–4070, 2013.

[155] A. J. Wunderlich. The Katsevich inversion formula for cone-beam computed tomography. Master's thesis, Oregon State University, 2006.

[156] F. Xu and K. Mueller. Accelerating popular tomographic reconstruction algorithms on commodity PC graphics hardware. *IEEE Trans. Nucl. Sci.*, 52:654–663, 2005.

[157] F. Xu, W. Xu, M. Jones, B. Keszthelyi, J. Sedat, D. Agard, and K. Mueller. On the efficiency of iterative ordered subset reconstruction algorithms for acceleration on GPUs. *Comput. Methods Programs Biomed.*, 98:261–270, 2010.

[158] X. L. Xu, J. S. Liow, and S. C. Strother. Iterative algebraic reconstruction algorithms for emission computed tomography: A unified framework and its application to positron emission tomography. *Medical Physics*, 20:1675–1684, 1993.

[159] J. Yang and Y. Zhang. Alternating direction algorithms for ℓ_1-problems in compressive sensing. *SIAM Journal on Scientific Computing*, 33:250–278, 2011.

[160] J. Yang, Y. Zhang, and W. Yin. A fast alternating direction method for TVL1-L2 signal reconstruction from partial fourier data. *IEEE Journal of Selected Topics in Signal Processing*, 4:288–297, 2010.

[161] W. Y. Yang and M. Slaney. *Signals and Systems with MATLAB*. Springer, 2009.

[162] W. Yin, S. Osher, D. Goldfarb, and J. Darbon. Bregman iterative algorithms for ℓ_1-minimization with applications to compressed sensing. *SIAM J. Imag. Sci.*, 1:143–168, 2008.

[163] H. Yu and G. Wang. A soft-threshold filtering approach for reconstruction from a limited number of projections. *Phys. Med. Biol.*, 55:3905–3916, 2010.

[164] 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 Med.*, 1:65–84, 2009.

[165] G. Zeng and G. Gullberg. A ray-driven backprojector for backprojection filtering and filtered backprojection algorithms. In *IEEE Nuclear Science Symposium Medical Imaging Conference*, pages 1199–1201, San Francisco, 1993.

[166] H. Zhang and W. W. Hager. A nonmonotone line search technique and its appli- cation to unconstrained optimization. *SIAM J. Optim.*, 14:1043–1056, 2004.

[167] W. Zhuang, S. S. Gopal, and T. J. Hebert. Numerical evaluation of methods for computing tomographic projections. *IEEE Transactions on Nuclear Science*, 41:1660–1665, 1994.

[168] A. Ziegler, T. Kohler, and R. Proksa. Noise and resolution in images reconstructed with FBP and OSC algorithms for CT. *Medical Physics*, 35:585–598, 2008.

[169] A. Ziegler, T. Nielsen, and M. Grass. Iterative reconstruction of a region of interest for transmission tomography. *Medical Physics*, 35:1317–1327, 2008.

Publications Arising from this Thesis

Some work described in this thesis has been published in the following papers:

1. Wen Hou and Cishen Zhang. Parallel-beam CT Reconstruction Based on Mojette Transform and Compressed Sensing. *International Journal of Computer and Electrical Engineering*, 2013, 5(1): 83-87.

2. Wen Hou and Cishen Zhang. A compressed sensing approach to low-radiation CT reconstruction. In *Proceedings of 9th IEEE/IET International Symposium on Communication Systems, Networks and Digital Signal Processing (CSNDSP)*, pages 793-797, Manchester, UK, 2014.

3. Wen Hou and Cishen Zhang. Analysis of Compressed Sensing Based CT Reconstruction with Low Radiation. In *Proceedings of International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)*, pages 291-296, Kuching, Sarawak, Malaysia, 2014. (Winner of the Best Paper Award in Signal Processing)