Utrecht University

Master Thesis

Reconstruction Algorithms for Spectral Computed Tomography

Author
Hongyan Liu

Supervisor
Dr. Tristan van Leeuwen

October 10, 2018

Contents

Chapter 1

Introduction

X-ray Computed Tomography (CT) is a conventional medical imaging modality which allows for imaging of internal structures in objects by using measurements of projection data from different angles. Hardware developments of CT systems in recent years, such as photon-counting detectors and fast kV switching, have made it possible to acquire polychromatic x-ray projection data. Collection of multi-energy sinogram can not only be used for imaging of attenuation coefficients at different energy bins, but it also enables the reconstruction of material basis images when mass attenuation coefficients for different elements are given. Sufficient use of the spectral information in the measurements, however, requires the development of new reconstruction methods.

The aim of this master thesis project is to develop and test fast and accurate algorithms for spectral CT reconstruction, and the thesis is organized in the following way.

Chapter 2 deals with single-energy reconstruction. To begin with, basic geometry and existing reconstruction methods are introduced for single-energy CT. Statistical reconstruction algorithms are then implemented and tested, with a focus on effects of regularization terms and acceleration techniques. Chapter 3 switches to multi-energy reconstruction, and a two-step algorithm is developed and experimented, after introduction of multi-energy physics and problem formulation for multi-energy reconstruction. Chapter 4 compares the algorithms for spectral CT reconstruction and draws conclusions, and presents an outlook for future research.

Chapter 2

Single-energy reconstruction

In this chapter, several existing reconstruction algorithms are introduced for classic single-energy CT technology. These algorithms will be basis steps for multi-energy CT reconstructions in later chapters. To begin with, Section 2.1 defines the geometry for the parallel CT reconstruction problem and introduces the filtered backprojection (FBP) method. Section 2.2 introduces two algebraic iterative reconstruction algorithms, ART and SART, and proves the equivalence between SART and gradient descent algorithm for a weighted least-squares problem. Section 2.3 explains how a maximum likelihood (ML) reconstruction model is developed considering single-energy photon statistics, and how the ML model can be approximated as a weighted least-squares problem. In Section 2.4, Tikhonov and TV regularization terms are introduced and numerical methods for solving Tikhonov- or TV-regularized statistical reconstruction problems are discussed. Algorithms discussed in Section 2.4 are tested finally by using simulation data, and results for single-energy reconstruction are summarized in Section 2.5 and provide inspiration for the algorithm design for multi-energy reconstruction in later chapters.

2.1 Filtered Backprojection(FBP)

Given a two-dimensional image $u(x,y) : \mathbb{R}^2 \to \mathbb{R}$, $p(\theta, b)$, the projection of u along a straight line $x\cos\theta + y\sin\theta = b$, can be expressed as a two dimensional integral equation

$$p(\theta,b) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} u(x,y)\delta(x\cos\theta + y\sin\theta - b)dxdy, \tag{2.1}$$

where $\theta \in [0, 2\pi)$, $b \in \mathbb{R}$, and $\delta(\cdot)$ is the Dirac delta function. The X-ray CT reconstruction problem in two dimensions is to recover $u(x,y)$ from projections $p(\theta,b)$ for all different θ. A projection $p(\cdot,\cdot) : \mathbb{R}^2 \to \mathbb{R}$ is called a sinogram because $p(\cdot,b)$ has a sine-like wave shape. The two-dimensional geometry of the parallel CT is illustrated in Figure 2.1.

Consider a rotated ξ–η coordinate system which has the following relationship with the original x–y coordinates,

$$\begin{cases} \xi = x\cos\theta + y\sin\theta, \\ \eta = -x\sin\theta + y\cos\theta, \end{cases} \tag{2.2}$$

and the projection integral (2.1) can be rewritten as

$$\begin{aligned} p(\theta,b) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} u\left(x(\xi,\eta,\theta), y(\xi,\eta,\theta)\right)\delta(\xi - b)d\xi d\eta \\ &= \int_{-\infty}^{\infty} u\left(x(b,\eta,\theta), y(b,\eta,\theta)\right)d\eta = \int_{-\infty}^{\infty} u\left(b\cos\theta - \eta\sin\theta, b\sin\theta + \eta\cos\theta\right)d\eta. \end{aligned} \tag{2.3}$$

The above equation shows that projection $p(\theta,b)$ can be computed by a line integral of the image $u(x,y)$. Conversely, in order to reconstruct the image from projections, the mostly widely

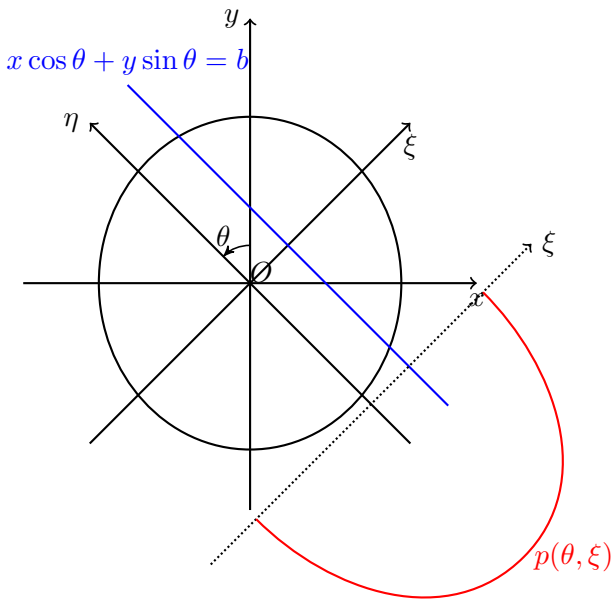


Figure 2.1: Geometry of 2D parallel CT. Rotating the x–y coordinate system counterclockwise by angle θ gives the new ξ–η coordinate system. Blue line: projection line $x\cos\theta + y\sin\theta = b$; Red curve: sinogram $p(\theta, \xi)$.

used algorithm is the classic filtered backprojection(FBP) method. The theorem which lays the foundation for FBP is called the Central Slice Theorem, which establishes a relationship between the one-dimensional Fourier transform of $p(\theta, b)$ and the two-dimension Fourier transform of $u(x, y)$. The two-dimension Fourier transform of $u(x, y)$ in $\theta - q$ polar coordinate system can be computed as

$$U(\theta, q) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} u(x, y) \exp\{-i2\pi(xq\cos\theta + yq\sin\theta)\}\, dxdy, \tag{2.4}$$

and the one-dimension Fourier transform of $p(\theta, b)$ can be computed as

$$\begin{aligned} P(\theta, q) &= \int_{-\infty}^{\infty} p(\theta, \xi) \exp\{-i2\pi q\xi\}\, d\xi \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} u(x(\xi, \eta, \theta), y(\xi, \eta, \theta)) \exp\{-i2\pi q\xi\}\, d\eta d\xi \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} u(x, y) \exp\{-i2\pi(xq\cos\theta + yq\sin\theta)\}\, dxdy = U(\theta, q). \end{aligned} \tag{2.5}$$

The second equality in (2.5) is obtained by substituting $p(\theta, \xi)$ by equation (2.3), and the third equality is obtained by coordinate transformation.

The Central Slice Theorem suggests that $u(x, y)$ can be reconstructed by computing the two-dimensional inverse Fourier transform of $P(\theta, q)$.To further simplify computation, the two-dimensional inverse transform can be rewritten as

$$\begin{aligned} u(x, y) &= \int_0^{2\pi}\int_0^{\infty} P(\theta, q) \exp\{i2\pi q(x\cos\theta + y\sin\theta)\}\, qdqd\theta \\ &= \int_0^{\pi}\int_{-\infty}^{\infty} P(\theta, q) \exp\{i2\pi q\xi\}\, |q|\, dqd\theta. \end{aligned} \tag{2.6}$$

The second equality in (2.6) uses the symmetry property of $P(\theta, q)$, since $P(\theta + \pi, q) = P(\theta, q)$ can be derived from $p(\theta + \pi, \xi) = p(\theta, \xi)$.Equation (2.6) suggests an algorithm to reconstruct

$u(x, y)$ by computing backprojection of the filtered sinogram without directly computing a two-dimensional inverse Fourier transform, and this is the filtered backprojection(FBP) algorithm, which is summarized in Algorithm 1.

Algorithm 1 Filtered Backprojection.

1: Compute the Fourier transform of $p(\theta, \xi)$:

$$p(\theta, \xi) \rightarrow P(\theta, q);$$

2: Compute the high-pass filtering of $P(\theta, q)$:

$$P(\theta, q) \rightarrow P(\theta, q) \left| q \right| ;$$

3: Compute the inverse Fourier transform of $P(\theta, q) \left| q \right|$:

$$P(\theta, q) \left| q \right| \rightarrow \int_{-\infty}^{\infty} P(\theta, q) \exp \left\{ i 2 \pi q \xi \right\} \left| q \right| dq \triangleq h(\theta, \xi);$$

4: Compute the backprojection on the line $\xi = x \cos \theta + y \sin \theta$:

$$u(x, y) = \int_{0}^{\pi} h(\theta, \xi) d\theta$$

2.2 Algebraic Reconstruction Technique(ART)

It is shown in the previous section that the relationship between $u(x, y)$ and $p(\theta, \xi)$ is linear. For real applications, projection measurements are discretized in θ and ξ directions, and the reconstructed image is discrete in x and y directions (see Figure 2.1). The relationship between a discretized image $\boldsymbol{u}$ and a set of discrete measurements $\boldsymbol{p}$ can therefore be written as

$$\boldsymbol{p} = \boldsymbol{W} \boldsymbol{u}, \tag{2.7}$$

where $\boldsymbol{u}$ and $\boldsymbol{p}$ are vectors of length m and n respectively, and $\boldsymbol{W}$ is a matrix of size $m \times n$ and is often called the system matrix. In typical CT scans, image $\boldsymbol{u}$ is the reconstructed image of attenuation coefficients. Various methods have been developed for computing matrix $\boldsymbol{W}$, and accurate methods such as the distance-driven method consider not only the image pixel size but the X-ray detector size as well [1]. The simplest method for computing $\boldsymbol{W}$ is to compute the intersection length between every pixel and every projection ray, as illustrated in Figure 2.2.

The earliest iterative method applied to solve $\boldsymbol{u}$ in (2.7) is called the Algebraic Reconstruction Technique [2]. The ART algorithm is given by Algorithm 2, in which $\boldsymbol{w}_i, i = 1, ...m$ are the row vectors of size $1 \times n$ in matrix $\boldsymbol{W}$. In ART, image updates $\boldsymbol{u}^{(0)}, \boldsymbol{u}^{(1)}, ..., \boldsymbol{u}^{(i)}, ...$ can be represented by vectors all starting from the origin in an n-dimension space, and by computing (2.8), an updated $\boldsymbol{u}^{(k)}$ is obtained by projecting $\boldsymbol{u}^{(k-1)}$ on the hyperplane represented by equation $\langle \boldsymbol{w}_i, \boldsymbol{u} \rangle - p^i = 0$. Convergence of this method can be shown geometrically in Figure 2.3 when an exact solution exists.

Instead of using projection data sequentially, an improved version of ART was developed by simultaneously using all projection data in one iteration [3]. This method is called Simultaneous algebraic reconstruction technique (SART), and the iteration step is given by

$$u_j^{(k)} = u_j^{(k-1)} - \frac{\displaystyle\sum_{i'=1}^{m} \frac{\langle \boldsymbol{w}_{i'}, \boldsymbol{u}^{(k-1)} \rangle - p_{i'}}{\sum_{j'=1}^{n} w_{i',j'}} \cdot w_{i',j}}{\sum_{i=1}^{m} w_{i,j}}. \tag{2.9}$$

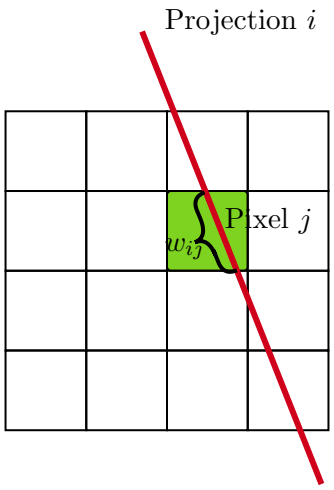


Figure 2.2: Illustration of computing w_{ij} in system matrix $\boldsymbol{W}$. w_{ij} can be computed as the intersection length between the ith projection line and the jth image pixel assuming a point x-ray source.

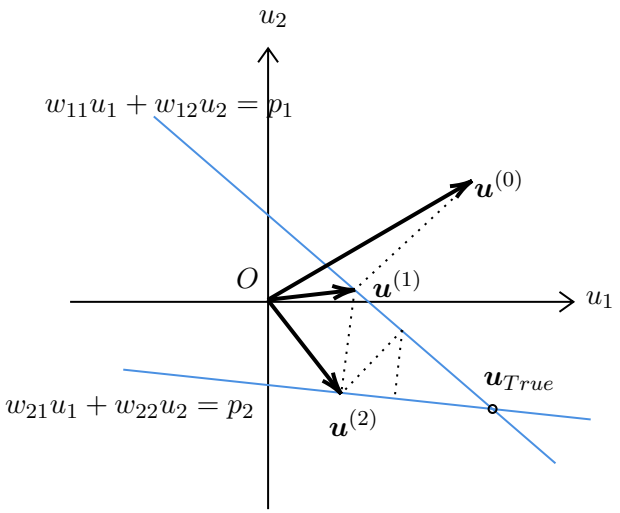


Figure 2.3: Geometrical illustration of ART algorithm 2 for solving a 2×2 system $\boldsymbol{W}\boldsymbol{u} = \boldsymbol{p}$. Every equation in the linear system is represented by a blue line in the two-dimensional space, and vectors $\boldsymbol{u}^{(0)}$, $\boldsymbol{u}^{(1)}$, ... are iterative updates for solving the system.

Algorithm 2 Algebraic Reconstruction Technique.

1: Start from an initial guess: $\boldsymbol{u}^{(0)}$;
2: **while** some stopping criterion is not met **do**
3: Select projection index $i \in [1, m]$ randomly;
4: Update the image:

$$\boldsymbol{u}^{(k)} = \boldsymbol{u}^{(k-1)} - \frac{\langle \boldsymbol{w}_i, \boldsymbol{u}^{(k-1)} \rangle - p_i}{\langle \boldsymbol{w}_i, \boldsymbol{w}_i \rangle} (\boldsymbol{w}_i)^T; \tag{2.8}$$

5: **end while**

The above equation can be rewritten in a compact form as

$$\boldsymbol{u^{(k)}} = \boldsymbol{u^{(k-1)}} - \boldsymbol{C}^{-1}\boldsymbol{W}^T\boldsymbol{R}^{-1}(\boldsymbol{W}\boldsymbol{u^{(k-1)}} - \boldsymbol{p}), \tag{2.10}$$

where $\boldsymbol{R}$ and $\boldsymbol{C}$ are diagonal matrices with diagonal elements $r_i = \sum_{j=1}^{n} w_{i,j}$ and $c_j = \sum_{i=1}^{m} w_{i,j}$ respectively. It can be observed that the SART iteratively solves a weighted least-squares problem expressed as

$$\boldsymbol{u}^* = \arg\min_{\boldsymbol{u}} f(\boldsymbol{u}) = \arg\min_{\boldsymbol{u}} \frac{1}{2} \|\boldsymbol{W}\boldsymbol{u} - \boldsymbol{p}\|^2_{R^{-1}}, \tag{2.11}$$

by a preconditioned gradient descent method with matrix $\boldsymbol{C}^{-1}$ only used for preconditioning. The gradient descent update with a step size β for problem (2.11) can be written as

$$\boldsymbol{u^{(k)}} = \boldsymbol{u^{(k-1)}} - \beta\boldsymbol{W}^T\boldsymbol{R}^{-1}(\boldsymbol{W}\boldsymbol{u^{(k-1)}} - \boldsymbol{p}), \tag{2.12}$$

and with sufficiently small step size, the gradient descent method ensures convergence for the quadratic problem (2.11). Difference between (2.10) and (2.12) is that equation (2.10) has parameter-dependent step sizes whereas equation (2.12) has a general step size.

The convergence of SART is proved below. Substitute $\boldsymbol{u}^{(k)}$ by equation (2.10), and the second-order Taylor expansion of $f(\boldsymbol{u}^{(k)})$ around $\boldsymbol{u}^{(k-1)}$ can be written as

$$\begin{aligned} f(\boldsymbol{u}^{(k)}) &= f(\boldsymbol{u}^{(k-1)}) + \nabla f(\boldsymbol{u}^{(k-1)})^T \cdot (\boldsymbol{u}^{(k)} - \boldsymbol{u}^{(k-1)}) + \frac{1}{2}(\boldsymbol{u}^{(k)} - \boldsymbol{u}^{(k-1)})^T \nabla^2 f(\boldsymbol{u}^{(k-1)})(\boldsymbol{u}^{(k)} - \boldsymbol{u}^{(k-1)}) \\ &= f(\boldsymbol{u}^{(k-1)}) - \nabla f(\boldsymbol{u}^{(k-1)})^T \boldsymbol{C}^{-1} \nabla f(\boldsymbol{u}^{(k-1)}) + \nabla f(\boldsymbol{u}^{(k-1)})^T \boldsymbol{C}^{-1}\boldsymbol{W}^T\boldsymbol{R}^{-1}\boldsymbol{W}\boldsymbol{C}^{-1} \nabla f(\boldsymbol{u}^{(k-1)}). \end{aligned} \tag{2.13}$$

To ensure $f(\boldsymbol{u}^{(k)}) < f(\boldsymbol{u}^{(k-1)})$, the following inequality

$$\nabla f(\boldsymbol{u}^{(k-1)})^T \left(\boldsymbol{C}^{-1} - \boldsymbol{C}^{-1}\boldsymbol{A}^T\boldsymbol{R}^{-1}\boldsymbol{A}\boldsymbol{C}^{-1}\right) \nabla f(\boldsymbol{u}^{(k-1)}) \geq 0 \tag{2.14}$$

should hold for any $\boldsymbol{u}^{(k-1)}$. Therefore, the preconditioned gradient descent method converges for arbitrary quadratic function f and preconditioning matrix $\boldsymbol{C}^{-1}$ as long as matrix $\boldsymbol{C}^{-1} - \boldsymbol{C}^{-1}\boldsymbol{A}^T\boldsymbol{R}^{-1}\boldsymbol{A}\boldsymbol{C}^{-1}$ is positive semi-definite. With $\boldsymbol{W}$ a system matrix with non-negative entries, $\boldsymbol{C}$ a diagonal matrix of column sums of $\boldsymbol{W}$ and $\boldsymbol{R}$ a diagonal matrix of row sums of $\boldsymbol{W}$, the positive semi-definite property can be proved by definitions of matrices $\boldsymbol{R}$ and $\boldsymbol{C}$ [4], therefore the convergence of SART is ensured. From the above analysis, it can be observed that the SART has linear convergence rate, and the preconditioning matrix $\boldsymbol{C}$ may be modified to further accelerate the convergence for this preconditioned gradient descent type method [4].

Instead of solving problem (2.11) by SART or other optimization methods, the reconstruction problem can also be formalized directly from model (2.7) as

$$\boldsymbol{u}^* = \arg\min_{\boldsymbol{u}} f(\boldsymbol{u}) = \arg\min_{\boldsymbol{u}} \frac{1}{2} \|\boldsymbol{W}\boldsymbol{u} - \boldsymbol{p}\|^2_2, \tag{2.15}$$

and quasi-Newton method can be used for solving this quadratic problem with a superlinear convergence rate in theory [5], which is faster than gradient descent based methods.

2.3 Statistical Reconstruction Technique (SRT)

Problems (2.11) and (2.15) formed in previous section have not taken into account the photon statistics. In order to obtain more accurate reconstruction results, a maximum likelihood method will be used and the problem will be further simplified as a weighted least-squares problem. Reconstruction models formalized considering the measurement statistics are called statistical reconstruction, as will be explained in the following.

Throughout this thesis, only the simplest Poission noise model is used [6]. For monochromatic measurement at fixed energy E, the x-ray incident photon number I_{in}, the measured photon number detected at the kth projection I_k and the expected photon number $\bar{I}_k$ follow the relationship

$$I_k = I_{in}e^{-p_k}, \bar{I}_k = I_{in}e^{-\langle \boldsymbol{w}_k, \boldsymbol{u}\rangle}, \tag{2.16}$$

where p_k and $\boldsymbol{w}_k$ use the same definition as in previous section. The measured photon number I_k follows a Poisson distribution with expected value $\bar{I}_k$. Therefore, the conditional probability of obtaining measurement results I_1,I_2,..., I_m given the expected photon number $\bar{I}_1$,$\bar{I}_2$,..., $\bar{I}_m$, can be expressed as

$$P\left(I_1, ..., I_m | \bar{I}_1, ..., \bar{I}_m\right) = \prod_{k=1}^{m} \frac{\bar{I}_k^{I_k} \cdot e^{-\bar{I}_k}}{I_k!}. \tag{2.17}$$

Taking logarithm of both sides in (2.17) and substituting $\bar{I}_k$ by (2.16) gives a log-likelihood function for image $\boldsymbol{u}$ after ignoring the constant part independent of $\boldsymbol{u}$

$$L\left(I_1, I_2, ..., I_m | \boldsymbol{u}\right) = \sum_{k=1}^{m} \left(-I_k \langle \boldsymbol{w}_k, \boldsymbol{u}\rangle - I_{in}e^{-\langle \boldsymbol{w}_k, \boldsymbol{u}\rangle}\right). \tag{2.18}$$

Therefore, the maximum likelihood method for reconstructing $\boldsymbol{u}$ can be modeled as

$$\boldsymbol{u}^* = \arg\min_{\boldsymbol{u}}\{-L\left(I_1, I_2, ..., I_m | \boldsymbol{u}\right)\} = \arg\min_{\boldsymbol{u}} \sum_{k=1}^{m} \left(I_k \langle \boldsymbol{w}_k, \boldsymbol{u}\rangle + I_{in}e^{-\langle \boldsymbol{w}_k, \boldsymbol{u}\rangle}\right). \tag{2.19}$$

Note that the above problem is difficult to solve because of the exponential components and the large number of measurements m. In order to simplify problem (2.19), the exponential component $e^{-\langle \boldsymbol{w}_k, \boldsymbol{u}\rangle}$ is approximated by second-order Taylor expansion around $\langle \boldsymbol{w}_k, \boldsymbol{u}\rangle = p_k$,

$$e^{-\langle \boldsymbol{w}_k, \boldsymbol{u}\rangle} \approx e^{-p_k} - e^{-p_k}(\langle \boldsymbol{w}_k, \boldsymbol{u}\rangle - p_k) + \frac{1}{2}e^{-p_k}(\langle \boldsymbol{w}_k, \boldsymbol{u}\rangle - p_k)^2. \tag{2.20}$$

After substituting (2.20) into (2.19), re-ordering the terms and using relationship (2.16), a weighted least-squares problem is obtained as

$$\begin{aligned} \boldsymbol{u}^* &= \arg\min_{\boldsymbol{u}} \sum_{k=1}^{m} \left(I_k \langle \boldsymbol{w}_k, \boldsymbol{u}\rangle + I_{in}e^{-p_k} - I_{in}e^{-p_k}(\langle \boldsymbol{w}_k, \boldsymbol{u}\rangle - p_k) + \frac{1}{2}I_{in}e^{-p_k}(\langle \boldsymbol{w}_k, \boldsymbol{u}\rangle - p_k)^2\right) \\ &= \arg\min_{\boldsymbol{u}} \sum_{k=1}^{m} \left(\frac{1}{2}I_k(\langle \boldsymbol{w}_k, \boldsymbol{u}\rangle - p_k)^2\right) = \arg\min_{\boldsymbol{u}} \frac{1}{2}\|\boldsymbol{W}\boldsymbol{u} - \boldsymbol{p}\|_{\Sigma}^2, \end{aligned} \tag{2.21}$$

where Σ is a diagonal matrix with entries $I_1, I_2, ..., I_m$. It can also be proved that Σ is the inverse of the covariance matrix for projection vector $\boldsymbol{p}$.

2.4 Regularization for iterative reconstruction

In the previous two subsections, tomography reconstruction problems are formalized as quadratic problems (2.15) and (2.21) with or without full consideration of measurement statistics. Regularization terms need to be added in at least the following situations: (1) when the reconstructed image size n is larger than the measurement size m, so there will be no unique solution for (2.15) and (2.21); (2) when measure locations are not well designed, e.g. projection angle θ doesn't span the full π range (limited-angle problem [7]), and this will result in an ill-conditioned system matrix $\boldsymbol{W}$; (3) when a clean image with little noise is required by using noisy measurement

data; (4) when detected photon numbers $I_1, I_2, ..., I_m$ typically vary greatly, so the reconstruction problem (2.21) will be very ill-posed because of the ill-conditioned Σ.

In this section, two different types of regularization terms will be added to problem (2.21), namely the Tikhonov regularization, and the Total Variation regularization, and numerical methods for solving the regularized problems will also be introduced.

2.4.1 Tikhonov regularization

A Tikhonov regularization term can be included in the statistical reconstruction problem (2.21), which can be expressed as

$$\boldsymbol{u}^* = \arg\min_{\boldsymbol{u}} \left(\frac{1}{2} \|\boldsymbol{W}\boldsymbol{u} - \boldsymbol{p}\|_{\Sigma}^2 + \alpha \|\boldsymbol{L}\boldsymbol{u}\|_2^2 \right) = \arg\min_{\boldsymbol{u}} \left(\frac{1}{2} \|\boldsymbol{W}_{\Sigma}\boldsymbol{u} - \boldsymbol{p}_{\Sigma}\|_2^2 + \alpha \|\boldsymbol{L}\boldsymbol{u}\|_2^2 \right), \quad \text{with } \boldsymbol{W}_{\Sigma} = \Sigma^{1/2}\boldsymbol{W}, \boldsymbol{p}_{\Sigma} = \Sigma^{1/2}\boldsymbol{p}, \tag{2.22}$$

where $\boldsymbol{L}$ is the regularization operator which will be a matrix for Tikhonov cases, and α is the regularization parameter. Let the reconstructed image be of size $\sqrt{n} \times \sqrt{n}$, and the image vector $\boldsymbol{u}$ be $(u_{1,1}, u_{1,2}, ...u_{1,\sqrt{n}}, u_{2,1}, u_{2,2}, ..., u_{\sqrt{n},1}, ..., u_{\sqrt{n},\sqrt{n}})^T$. A typical smoothing operator $\boldsymbol{L}$ for image vector $\boldsymbol{u}$ can be defined as

$$\|\boldsymbol{L}\boldsymbol{u}\|_2^2 = \sum_{i=1}^{\sqrt{n}} \sum_{j=1}^{\sqrt{n}-1} (u_{i,j} - u_{i,j+1})^2 + \sum_{i=1}^{\sqrt{n}-1} \sum_{j=1}^{\sqrt{n}} (u_{i,j} - u_{i+1,j})^2, \tag{2.23}$$

which uses the finite difference method to approximate directional derivatives [8]. $\boldsymbol{L}$ defined by above (2.23) can also be written in a compact matrix form. Tikhonov regularized problem (2.22) has a closed form solution

$$\boldsymbol{u}^* = \left(\boldsymbol{W}^T\Sigma\boldsymbol{W} + 2\alpha\boldsymbol{L}^T\boldsymbol{L}\right)^{-1} \boldsymbol{W}^T\Sigma\boldsymbol{p}, \tag{2.24}$$

when $\boldsymbol{L}$ is properly designed so that matrix $\boldsymbol{W}^T\Sigma\boldsymbol{W} + 2\alpha\boldsymbol{L}^T\boldsymbol{L}$ is invertible.

The Tikhonov regularized problem (2.22) can be rewritten in a simple quadratic form

$$\boldsymbol{u}^* = \arg\min_{\boldsymbol{u}} \|\widetilde{\boldsymbol{W}}\boldsymbol{u} - \widetilde{\boldsymbol{p}}\|_2^2, \text{where } \widetilde{\boldsymbol{W}} = \begin{pmatrix} \Sigma^{1/2}\boldsymbol{W} \\ \sqrt{2\alpha}\boldsymbol{L} \end{pmatrix}, \widetilde{\boldsymbol{p}} = \begin{pmatrix} \Sigma^{1/2}\boldsymbol{p} \\ 0 \end{pmatrix}. \tag{2.25}$$

Numerical methods

A solution for the above least-squares problem (2.25) satisfies the normal equation $\widetilde{\boldsymbol{W}}^T\widetilde{\boldsymbol{W}}\boldsymbol{u} = \widetilde{\boldsymbol{W}}^T\widetilde{\boldsymbol{p}}$, which can be solved by numerical linear algebra methods such as conjugate gradient (CG). The solution for (2.25) is unique if matrix $\widetilde{\boldsymbol{W}}$ has full column rank. More generally, limited-memory quasi-Newton methods, such as the L-BFGS (limited-memory Broyden–Fletcher–Goldfarb–Shanno) method [9, 10], can be used for solving any differentiable convex minimization problem, and therefore is capable of solving the Tikhonov regularized problem (2.22), and has similar convergence rate comparing to conjugate gradient method [11]. An explicitly expressed L-BFGS method is summarized in Algorithm 3. Note that equation (2.27) gives an iterative expression for approximating the Hessian inverse, however in practical algorithms, $\boldsymbol{H}_k$ is never explicitly computed because of the large size of the matrix. By using the iterative expression (2.27), the matrix-vector multiplication $\boldsymbol{H}_k \nabla f(\boldsymbol{u}^{(k)})$ can be recursively computed by vector-vector multiplications and vector summations, so there is no need to explicitly store the large dense matrix H_k.

Besides the selection of an efficient numerical optimization algorithm such as L-BFGS, there are at least the following three aspects to consider for further convergence acceleration. Firstly,

Algorithm 3 L-BFGS quasi-Newton method for solving $\boldsymbol{u}^* = \arg\min_{\boldsymbol{u}\in\mathbb{R}^n} f(\boldsymbol{u})$.

1: Start from an initial guess: $\boldsymbol{u}^{(0)}$;
2: Initial estimate for Hessian inverse matrix: $k = 0, \boldsymbol{H}_0 = I_n$;
3: **while** $\|\nabla f(u^{(k)})\|^2 > tol$ **do**
4: Compute search direction: $\boldsymbol{d}_k = -\boldsymbol{H}_k \nabla f(\boldsymbol{u}^{(k)})$;
5: Search step size β_k such that $f(\boldsymbol{u}^{(k)}+\beta_k\boldsymbol{d}_k) < f(\boldsymbol{u}^{(k)})$; Specifically for quadratic problem (2.25), β_k can be computed by steepest descent:

$$\beta_k = \frac{\nabla f(\boldsymbol{u}^{(k)})^T \boldsymbol{H}_k \nabla f(\boldsymbol{u}^{(k)})}{2\boldsymbol{d}_k^T \widetilde{\boldsymbol{W}}^T \widetilde{\boldsymbol{W}} \boldsymbol{d}_k}; \tag{2.26}$$

6: Update $\boldsymbol{u}$: $\boldsymbol{u}^{(k+1)} = \boldsymbol{u}^{(k)} + \beta_k \boldsymbol{d}_k$;
7: Update $\boldsymbol{H}_{k+1}$: Start from $\boldsymbol{H}_{k+1} = I_n$;
8: **for** $j = \max(0, k - N_{memory} + 1) : k$ **do**

$$\boldsymbol{H}_{k+1} \leftarrow \left(I_n - \frac{\boldsymbol{p}_j \boldsymbol{q}_j^T}{\boldsymbol{p}_j^T \boldsymbol{q}_j}\right) \boldsymbol{H}_{k+1} \left(I_n - \frac{\boldsymbol{q}_j \boldsymbol{p}_j^T}{\boldsymbol{p}_j^T \boldsymbol{q}_j}\right) + \frac{\boldsymbol{p}_j \boldsymbol{p}_j^T}{\boldsymbol{p}_j^T \boldsymbol{q}_j}, \tag{2.27}$$

where $\boldsymbol{p}_j = \boldsymbol{u}^{(j+1)} - \boldsymbol{u}^{(j)}, \boldsymbol{q}_j = \nabla f(\boldsymbol{u}^{(j+1)}) - \nabla f(\boldsymbol{u}^{(j)})$.

9: **end for**
10: $k \leftarrow k + 1$;
11: **end while**

choosing a proper initial guess $\boldsymbol{u}^{(0)}$ near the true solution $\boldsymbol{u}^*$could greatly reduce the number of iterations required for Newton-type methods [12]. A smoothed FBP result can be a proper choice for $\boldsymbol{u}^{(0)}$.

Secondly, restart of the Hessian-inverse matrix approximation sometimes will be required to ensure a fast convergence for Algorithm 3. For Newton-type optimization algorithms, a fast convergence requires the inner product of search direction and gradient direction to be negative, which means $\nabla f(\boldsymbol{u}_k)^T \boldsymbol{d}_k < 0$. For quasi-Newton algorithms, this condition cannot be fulfilled when the estimated Hessian inverse matrix is very ill-conditioned. In Algorithm 3, matrix $\boldsymbol{H}_k$ is estimated by pairs of vectors from previous k iterations in the beginning N_{memory} iterations, and will be estimated by vectors from previous N_{memory} iterations after N_{memory} iterations. A larger memory size N_{memory} doesn't ensure a more accurate estimate nor a faster convergence. To ensure an overall fast convergence, a proper memory size N_{memory} needs to be selected and sometimes a restart of the algorithm will be required periodically every $N_{restart}$ iterations.

Finally, using a preconditioning technique can further accelerate the convergence for quadratic problem (2.25). Matrix $\widetilde{\boldsymbol{W}}$ typically has a very large conditioning number because of the large dynamic range of the statistical weights in matrix Σ, and this dramatically slows down the convergence of (2.22) compared to solving a uniformly weighted problem by choosing Σ as an identity matrix. When there exists a relationship $\boldsymbol{u} = \boldsymbol{P}\boldsymbol{y}$, where $\boldsymbol{y}$ is a vector of the same size as $\boldsymbol{u}$, and $\boldsymbol{P}$ is a full rank square matrix, problem (2.25) is equivalent to the following

$$\boldsymbol{y}^* = \arg\min_{\boldsymbol{y}} \|(\widetilde{\boldsymbol{W}}\boldsymbol{P})\boldsymbol{y} - \widetilde{\boldsymbol{p}}\|_2^2, \text{where } \boldsymbol{u} = \boldsymbol{P}\boldsymbol{y}. \tag{2.28}$$

When a proper preconditioning matrix $\boldsymbol{P}$ is designed such that the conditioning number of $\widetilde{\boldsymbol{W}}\boldsymbol{P}$ is smaller than that of the original matrix $\widetilde{\boldsymbol{W}}$, solving problem (2.28) will yield a faster convergence rate than solving the original problem (2.25). Similar as in [13], preconditioning matrix $\boldsymbol{P}$ for (2.22) can be designed as

$$\boldsymbol{P} = \operatorname{diag}\left\{\boldsymbol{W}_\Sigma^T \boldsymbol{W}_\Sigma \boldsymbol{1} + 2\alpha \boldsymbol{L}^T \boldsymbol{L}\boldsymbol{1}\right\}^{-1/2}, \tag{2.29}$$

where $\mathbf{1}$ is a vector of ones. The purpose for this preconditioning matrix is to balance the row sums of the original matrix $\boldsymbol{W}_\Sigma$.

Choosing regularization parameter

In order to obtain a reconstructed image in good quality for the Tikhonov regularized problem, it is important to choose a proper regularization parameter α. Regularization parameter balances the data fidelity term and the regularization term in the objective function. In general, two different methods can be used for choosing the regularization parameter. The first is called the L-curve method [14], in which the L-curve is defined as $(\|\boldsymbol{L}\boldsymbol{u}_\alpha\|_2, \|\boldsymbol{W}_\Sigma\boldsymbol{u}_\alpha - \boldsymbol{p}_\Sigma\|_2)_{\alpha\in(0,\infty)}$. Here, $\boldsymbol{u}_\alpha$ is the solution by using regularization parameter α, and an optimal α can be found by seeking the point on the L-curve with maximum curvature. In practice, the L-curve can be obtained by either deriving an analytic function or sampling multiple α and computing the corresponding $\boldsymbol{u}_\alpha$.

The second method for α selection is called Generalized Cross-Validation (GCV) [15], which minimizes the GCV function for problem (2.22)

$$GCV(\alpha) = \frac{\|\boldsymbol{W}_\Sigma\boldsymbol{u}_\alpha - \boldsymbol{p}_\Sigma\|_2^2}{\left(\text{trace}(I_m - \boldsymbol{W}_\Sigma\boldsymbol{W}_{\Sigma,\alpha}^\dagger)\right)^2}, \tag{2.30}$$

where $\boldsymbol{W}_{\Sigma,\alpha}^\dagger = \left(\boldsymbol{W}^T\Sigma\boldsymbol{W} + 2\alpha\boldsymbol{L}^T\boldsymbol{L}\right)^{-1}\boldsymbol{W}^T\Sigma$ is a 'regularized inverse'. Note that directly computing the trace of $\boldsymbol{W}_\Sigma\boldsymbol{W}_{\Sigma,\alpha}^\dagger$ is computationally expensive for large system matrix $\boldsymbol{W}$. Therefore one way for stochastically approximating the trace of a matrix called Hutchinson's estimator [16] is used, as illustrated in the following theorem.

Theorem 1 *(Stochastic trace estimator [16]) Given any matrix $\boldsymbol{A}$ of size $n\times n$, after generating a sequence of random vectors $\boldsymbol{v}$ which satisfy $\mathbb{E}(\boldsymbol{v}\boldsymbol{v}^T) = \boldsymbol{I}_n$, the trace of $\boldsymbol{A}$ can be estimated by using the equality $tr(\boldsymbol{A}) = \frac{1}{n}\mathbb{E}(\boldsymbol{v}^T\boldsymbol{A}\boldsymbol{v})$.*

2.4.2 Total Variation regularization

A discretized anisotropic total variation (TV) [17] based reconstruction can be expressed as

$$\begin{aligned}\boldsymbol{u}^* &= \arg\min_{\boldsymbol{u}}\{J_{err}(\boldsymbol{u}) + \alpha J_{ani}(\boldsymbol{u})\}\\ &= \arg\min_{\boldsymbol{u}}\left\{\frac{1}{2}\|\boldsymbol{W}\boldsymbol{u} - \boldsymbol{p}\|_\Sigma^2 + \alpha\left(\sum_{i=1}^{\sqrt{n}}\sum_{j=1}^{\sqrt{n}-1}|u_{i,j} - u_{i,j+1}| + \sum_{i=1}^{\sqrt{n}-1}\sum_{j=1}^{\sqrt{n}}|u_{i,j} - u_{i+1,j}|\right)\right\},\end{aligned} \tag{2.31}$$

where $J_{err}(\boldsymbol{u})$ is the data fidelity term and $J_{ani}(\boldsymbol{u})$ is the anisotropic TV regularization term. For such a non-differentiable convex problem, quasi-Newton methods as discussed in the previous section cannot be directly used. Various methods have been studied to solve these ROF (Rudin, Osher and Fatemi [18]) model based reconstruction problems, such as the first-order primal-dual algorithms [19, 20].

Problem (2.31) can be reformulated as

$$\boldsymbol{u}^*, \boldsymbol{g}^* = \arg\min_{\boldsymbol{u},\boldsymbol{g}}\{J_{err}(\boldsymbol{u}) + \alpha J_{L1}(\boldsymbol{g})\}, \text{ subject to } \boldsymbol{g} = \boldsymbol{L}_{div}\boldsymbol{u}, \tag{2.32}$$

where function $J_{L1}(\boldsymbol{g})$ computes the absolute entry sum of vector $\boldsymbol{g}$, which can be expressed as $J_{L1}(\boldsymbol{g}) = \sum_{k=1}^{N_g}|g_k|$, and directional gradient vector $\boldsymbol{g}$ yields a linear relationship with image vector $\boldsymbol{u}$, which can be expressed as

$$\boldsymbol{g} = \boldsymbol{L}_{div}\boldsymbol{u} = \left(\{u_{i,j} - u_{i,j+1}\}_{i=1,\ldots\sqrt{n},j=1,\ldots\sqrt{n}-1}, \{u_{i,j} - u_{i+1,j}\}_{i=1,\ldots\sqrt{n}-1,j=1,\ldots\sqrt{n}}\right)^T. \tag{2.33}$$

For reformulated problem (2.32), ADMM (Alternating Direction Method of Multipliers) algorithm [21] , which is closely related to the above mentioned primal-dual algorithms, can be applied to split the problem into several easily solved sub-problems. An ADMM algorithm for solving problem (2.32) is summarized in Algorithm 4. For the $\boldsymbol{u}_k$ update in the algorithm, sub-problem (2.34) is quadratic and almost the same as problem (2.22), therefore it can be approximately solved by quasi-Newton methods within a few iterations. For the $\boldsymbol{g}$ update, the non-smooth sub-problem (2.35) can be solved by element-wise soft thresholding. ρ in the algorithm is the augmented Lagrangian parameter, which influences the convergence rate. A similar ADMM algorithm for single-energy CT reconstruction has been developed in [22]. For splitting-based algorithms, the main idea is to separate the main problem into a few sub-problems which are all easy to solve, and especially for TV-regularized problem, the splitted non-smooth L1 sub-problem should be easy to solve by applying soft-thresholding.

Algorithm 4 ADMM algorithm for solving (2.32).

1: Choose initial guesses: $\boldsymbol{u}^{(0)}$, $\boldsymbol{g}^{(0)}$, $\boldsymbol{r}^{(0)}$;
2: **for** $k = 1, 2, 3, \ldots$ **do**

$$\boldsymbol{u}_{k+1} := \arg\min_{\boldsymbol{u}} \left(J_{err}(\boldsymbol{u}) + \frac{\rho}{2} \|\boldsymbol{L}_{div}\boldsymbol{u} - \boldsymbol{g}_k + \boldsymbol{r}_k\|_2^2 \right) \tag{2.34}$$

$$\boldsymbol{g}_{k+1} := \arg\min_{\boldsymbol{g}} \left(J_{L1}(\boldsymbol{g}) + \frac{\rho}{2} \|\boldsymbol{L}_{div}\boldsymbol{u}_{k+1} - \boldsymbol{g} + \boldsymbol{r}_k\|_2^2 \right) \tag{2.35}$$

$$\boldsymbol{r}_{k+1} := \boldsymbol{r}_k + \boldsymbol{L}_{div}\boldsymbol{u}_{k+1} - \boldsymbol{g}_{k+1} \tag{2.36}$$

3: **end for**

Another way to solve the TV-based reconstruction is to approximate the non-smooth problem by the following differentiable convex problem

$$\begin{aligned} \boldsymbol{u}^* &= \arg\min_{\boldsymbol{u}} \left\{ J_{err}(\boldsymbol{u}) + \alpha \widetilde{J}_{ani}(\boldsymbol{u}) \right\} \\ &= \arg\min_{\boldsymbol{u}} \left\{ \frac{1}{2}\|\boldsymbol{W}\boldsymbol{u} - \boldsymbol{p}\|_{\Sigma}^2 + \alpha \left(\sum_{i=1}^{\sqrt{n}} \sum_{j=1}^{\sqrt{n}-1} \sqrt{|u_{i,j} - u_{i,j+1}|^2 + \epsilon} + \sum_{i=1}^{\sqrt{n}-1} \sum_{j=1}^{\sqrt{n}} \sqrt{|u_{i,j} - u_{i+1,j}|^2 + \epsilon} \right) \right\}, \end{aligned} \tag{2.37}$$

where function $\widetilde{J}_{ani}$ is the approximated TV term, and ϵ is a smoothing parameter with small positive value. Such a 'corner-rounding' differentiable approximation approach has been used in solving MRI reconstruction problems [23]. Approximated problem (2.37) can now be solved by the L-BFGS Algorithm 3 discussed in the previous subsection. Specifically when solving the non-quadratic problem (2.37), the step size α_k can not be computed by equation (2.26) but can be determined by a backtracking line search. After choosing a proper initial guess, a well-designed preconditioning matrix and a proper restart scheme, the approximated problem (2.37) can be solved by L-BFGS algorithm with a similar convergence rate as the Tikhonov problem, and the result of (2.37) will be very similar to the result of the original problem (2.31) if the smoothing parameter ϵ is small enough.

2.5 Numerical experiments

2.5.1 Experimental setups

Modified Shepp–Logan phantom

The Shepp-Logan phantom was firstly designed in 1974 for validating image reconstruction algorithms [24]. In order to test algorithms for reconstructing attenuation coefficients of a head

Body tissues	Mass attenuation coefficients μ/ρ (m^2kg^{-1})	The density of tissue (kg m^{-3}) in 20 °C	Attenuation coefficients $\mu(mm^{-1})$
Skeleton-cortical bone	2.23×10^{-2}	1920	0.0428
Water	1.84×10^{-2}	1000	0.0184
Brain (grey/white matter 50:50)	1.83×10^{-2}	1040	0.0190
Blood (whole)	1.83×10^{-2}	1060	0.0194

Table 2.1: Attenuation coefficients and the related densities for different tissues at 80 keV.

model, a modified Shepp-Logan phantom is required here. Four different body tissues are set in the phantom, namely cortical bone, water, brain and blood. Table 2.1 lists the attenuation coefficients for different tissues at 80keV as provided by ICRU 44 report [25], and is used for designing the modified phantom for simulation studies in this chapter. The dimensions of the phantom are 300×300 mm^2, consisting of 256×256 pixels, and the true modified phantom used here is shown in Figure 2.4a.

(a) Modified Shepp–Logan phantom. (b) Smoothed FBP (290 projections). (c) Smoothed FBP (58 projections).

Figure 2.4: True Phantom and FBP reconstruction results. All images are displayed with window $[0.018, 0.020]mm^{-1}$. Smoothed FBP results are obtained by applying a Hann window to the frequency domain sinogram, and a frequency scaling parameter is used for modifying the Hann filter by rescaling the frequency axis. For the 290-projection experiment, the frequency scaling parameter is set to 0.20, and for the 58-projection experiment it is set to 0.15.

CT geometry and system matrix $\boldsymbol{W}$

For simplicity, a two-dimensional parallel-beam CT imaging geometry is used for experiments throughout the thesis. There are 672 equally spaced bins on a 300mm-long one-dimensional detector, thus the distance between the centers of two neighbouring detector bins will be $300/672 \approx 0.446$ mm. For the number of projections or views measured in the 2π rotating range, there are two different conditions considered in the experiments in this chapter. 290 views are measured for a sufficient view problem and 58 views are measured for an insufficient view or compressed sensing problem.

The system matrix or projection matrix $\boldsymbol{W}$ models the relationship between the discrete image and the noise-free measurements. It will be a large sparse matrix. For example, for the sufficient view problem the size of $\boldsymbol{W}$ will be $(290 * 672) \times (256 * 256)$. In this thesis, the

	Sufficient view	Insufficient view
Number of projections	290	58
Incident photon number I_{in}	$2.5 * 10^5$	$1.25 * 10^6$

Table 2.2: Two different measurement conditions, sufficient and insufficient view.

ASTRA (All Scale Tomographic Reconstruction Antwerp) toolbox [26, 27] is used for generating the $\boldsymbol{W}$. The ASTRA toolbox is an open platform for image reconstruction in tomography, and it provides flexible functions to set up image and projection geometries and the corresponding projection matrix. It also provides various existing algorithms to conduct image reconstructions.

Generation of simulated noisy sinogram

The noise-free sinogram $\boldsymbol{p}_{clean}$ represents the projected attenuation coefficient measurements, and can be computed directly given the projection matrix $\boldsymbol{W}$ and the true image $\boldsymbol{u}$ simulated as in previous subsections. Algorithm 5 describes the process to generate a noisy sinogram using the Possion noise model as described in the previous subsection 2.3. Note that the simulated attenuation coefficients will first be converted to photon numbers in order to apply the Possion noise model, and the noisy photon number measurements will then be converted back. For simplicity, in experiments throughout this thesis, x-ray source photon number I_{in} at a given energy E will be set to a constant for all detector bins. However in real experiments, a wedge-shaped x-ray beam filter is usually used to improve image quality and in such cases I_{in} will be different for different detectors [28].

For experiments in this chapter, input x-ray photon number I_{in} will be set to $2.5 * 10^5$ for 290-view experiments, and $1.25 * 10^6$ for 58-view experiments. This design is to ensure that the total energy exposed remains the same for different experimental setups, and the two different measurement conditions are summarized in Table 2.2. For comparison with the iterative reconstruction results, smoothed FBP images are shown in Figure 2.4b and 2.4c. The MATLAB function *iradon* is used for obtaining the Hann-filtered results. It can be seen that the FBP few-view reconstruction result becomes much worse compared to the sufficient-view reconstruction, even though the total incident photon number for the two measurement conditions is the same.

Algorithm 5 Generate noisy sinogram $\boldsymbol{p}$.

1: Input: I_{in}, x-ray source photon number; $\boldsymbol{p}_{clean}$, simulated clean sinogram;
2: Compute expected photon number:

$$\bar{I}_k = I_{in} \cdot e^{-p_{clean,k}}, \text{ for } k = 1, 2, ..., m;$$

3: Compute measured photon number with Poisson noise:

$$I_k = \text{Poisson}(\bar{I}_k), \text{ for } k = 1, 2, ..., m;$$

4: Compute noisy sinogram:

$$p_k = \ln I_{in} - \ln I_k, \text{ for } k = 1, 2, ..., m;$$

Reconstruction model	Algorithm	Convergence acceleration
Tikhonov-regularized model (2.22)/(2.25)	Quasi-Newton (see Algorithm 3)	Preconditioning
TV-regularized model (2.31)	ADMM (See Algorithm 4) for solving splitted TV model (2.32)	Tuning augmented Lagrangian parameter
	Quasi-Newton for solving smoothed TV model (2.37)	periodic algorithm restart(1), better initial guess(2), circular ROI(3), Preconditioning(4)...

Table 2.3: Models and algorithms for single-energy experiments.

Tested models and algorithms

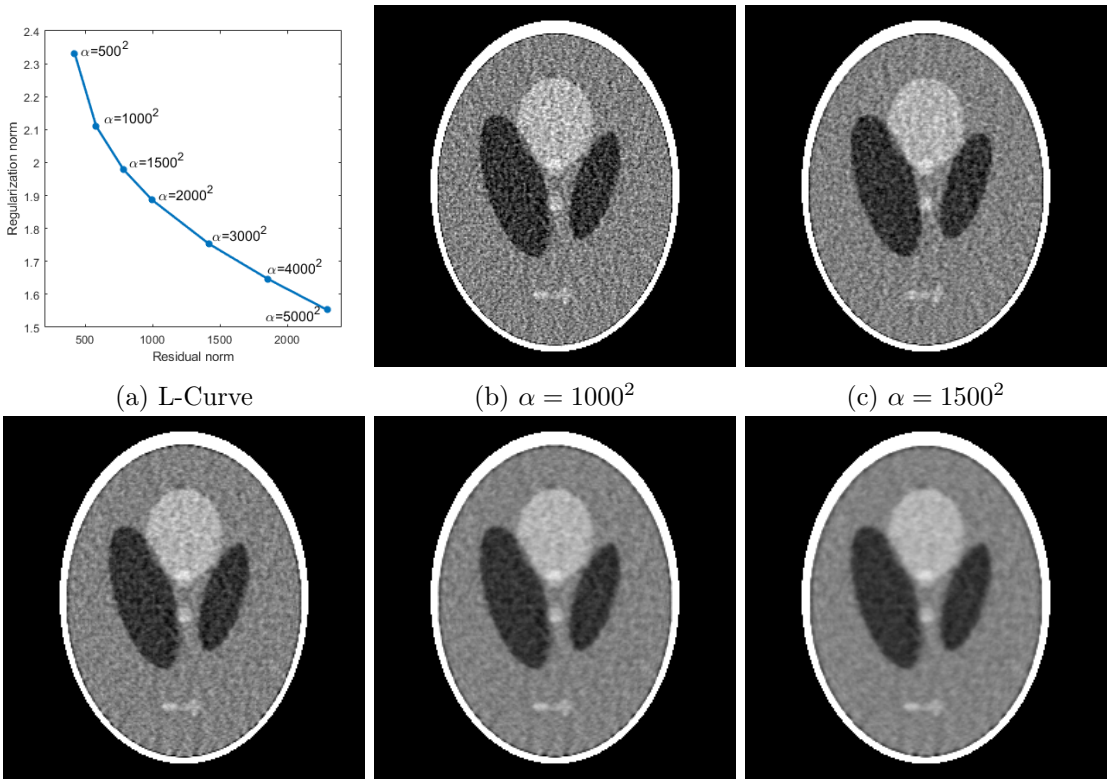
Two reconstruction models with different regularization terms and corresponding algorithms used for solving them are summarized in Table 2.3. More theoretical details have been given in the previous Section 2.4. For both experimental models and all the algorithms, three different aspects are tested and summarized in the following subsection 2.5.2, firstly the regularization parameter selection, secondly the qualitative and quantitative evaluation of the reconstructed images, and finally the acceleration techniques that can be used and convergence analysis of the algorithms.

2.5.2 Results

Regularization parameter selection

For both model (2.22) and (2.31), regularization parameter α plays a significant role in balancing image resolution and noise. Determining a proper α value would be a first step for all regularized-model reconstructions. In this chapter, the L-curve method as discussed in Section 2.4.1 is used for both Tikhonov model (2.22) and TV model (2.31) [29] because of its simplicity in computation. Since the selection of different algorithms and different acceleration techniques doesn't influence the final image solution and the α selection result, the quasi-Newton algorithm is used for computing all the Tikhonov model and smoothed TV model reconstructions, and no further convergence acceleration technique is applied here. To ensure convergence to a stable solution, the algorithm is run up to 200 iteration steps for every Tikhonov model experiment, and is run up to 500 iteration steps with a periodical restart step size of 20 for every TV model experiment.

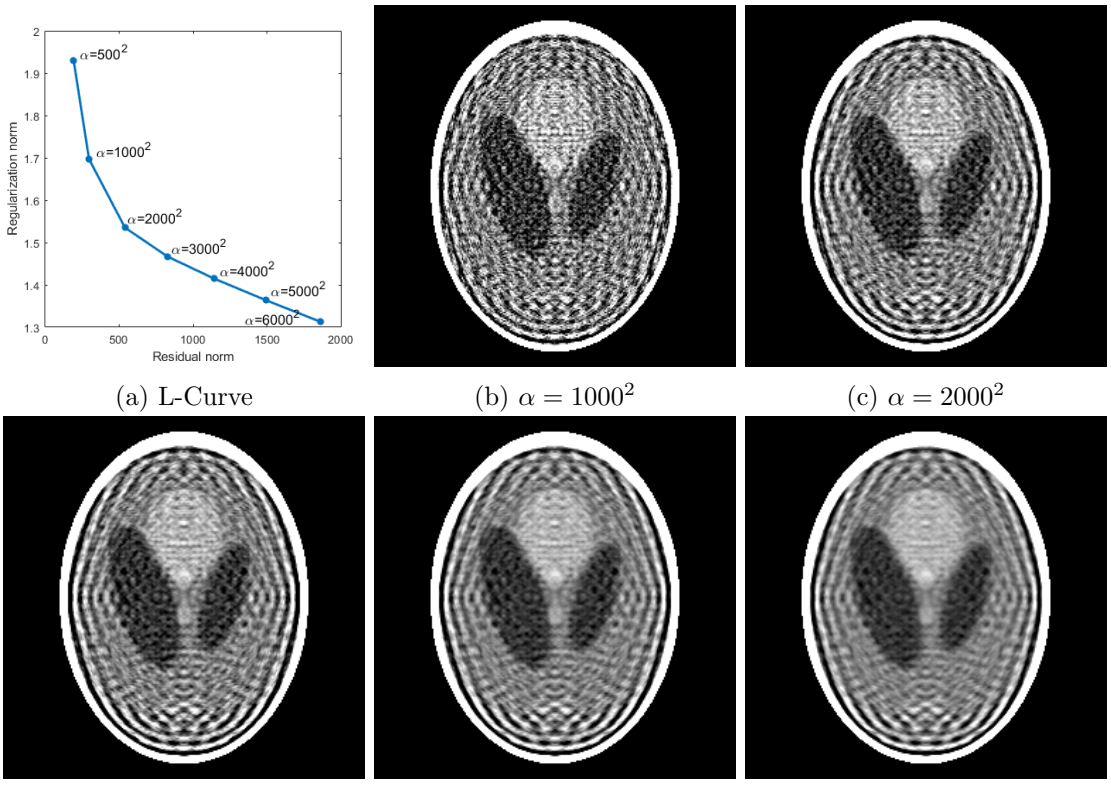
The L-curve for both regularized-models and measurement conditions is plotted in Figure 2.5a, 2.6a, 2.7a and 2.8a, and the corresponding reconstruction images are shown in Figure 2.5b-2.5f, 2.6b-2.6f, 2.7b-2.7f, and 2.8b-2.8f. In all the four discrete L-curve plots, the curvature change within a small range of α is not obvious. Therefore, it is difficult to find the corner point and thus decide an accurate regularization parameter only based on the L-curve plot. α values for all the four experiment settings are estimated from the L-curve results and summarized in Table 2.4, and choosing different α values in a reasonable range leads to reconstruction images with different qualities (such as noise level and resolution). A comparison of the reconstruction results shown in Figure 2.5-2.8 demonstrates that the TV model provides relatively good reconstruction results in both sufficient and insufficient projection settings, whereas the Tikhonov model could not suppress the streak artifacts caused by insufficient angular sampling and thus provides very poor reconstruction results (Figure 2.6b - 2.6f) in insufficient projection settings.



(a) L-Curve (b) $\alpha = 1000^2$ (c) $\alpha = 1500^2$

(d) $\alpha = 2000^2$ (e) $\alpha = 3000^2$ (f) $\alpha = 4000^2$

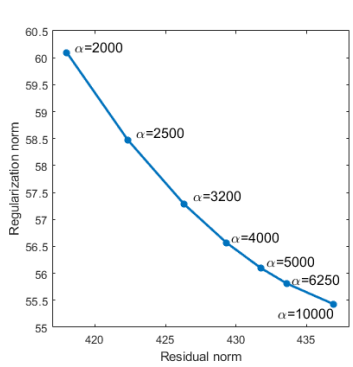
Figure 2.5: L-Curve plot and image results for various α. Tikhonov-regularized model, 290 projections. Display window $[0.018, 0.020]\text{mm}^{-1}$.



(a) L-Curve (b) $\alpha = 1000^2$ (c) $\alpha = 2000^2$

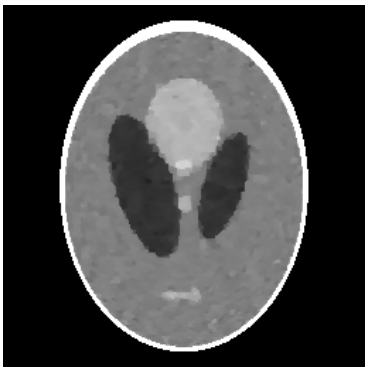
(d) $\alpha = 3000^2$ (e) $\alpha = 4000^2$ (f) $\alpha = 5000^2$

Figure 2.6: L-Curve plot and image results for various α. Tikhonov-regularized model, 58 projections. Display window $[0.018, 0.020]\text{mm}^{-1}$.

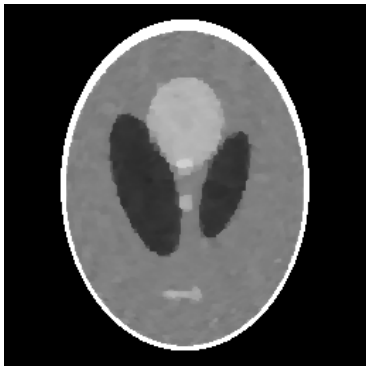


(a) L-Curve

(b) $\alpha = 2000$



(c) $\alpha = 3200$

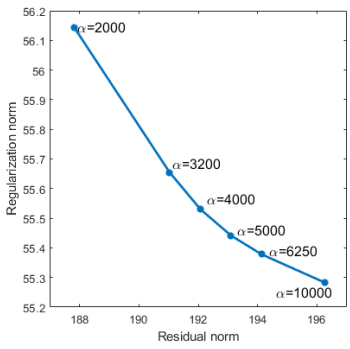


(d) $\alpha = 4000$

(e) $\alpha = 5000$

(f) $\alpha = 10000$

Figure 2.7: L-Curve plot and image results for various α. TV-regularized model, 290 projections. Display window $[0.018, 0.020]\text{mm}^{-1}$.



(a) L-Curve

(b) $\alpha = 2000$

(c) $\alpha = 3200$

(d) $\alpha = 4000$

(e) $\alpha = 5000$

(f) $\alpha = 10000$

Figure 2.8: L-Curve plot and image results for various α. TV-regularized model, 58 projections. Display window $[0.018, 0.020]\text{mm}^{-1}$.

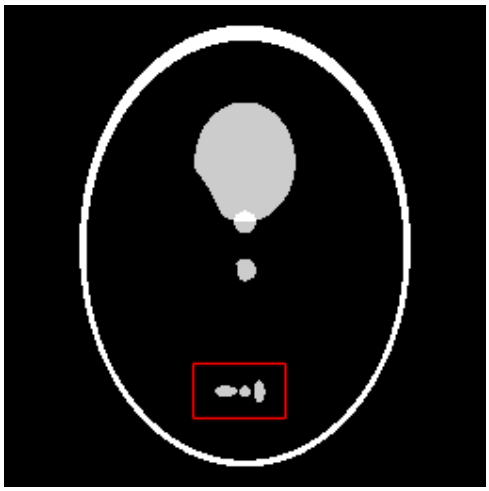
	Tikhonov		TV	
	290 views	58 views	290 views	58 views
α	$1000^2 - 3000^2$	$1000^2 - 3000^2$	$3200 - 10000$	$3200 - 10000$

Table 2.4: Regularization parameterα selection for L-curve for both models and measurement conditions.

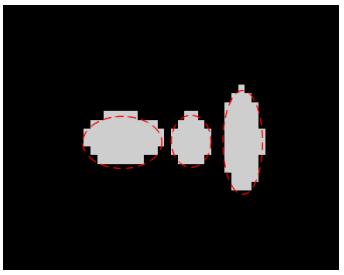
Qualitative and quantitative evaluation

In the previous paragraph, it is shown that 3 of all the 4 experiments, 290-projection-Tikhonov, 290-projection-TV, 58-projection-TV, have acceptable reconstruction results. To further evaluate and compare the results in the 3 experiments, more detailed visual inspections of the reconstructed images are conducted, and more quantitative metrics are applied for analysis in the following paragraphs.

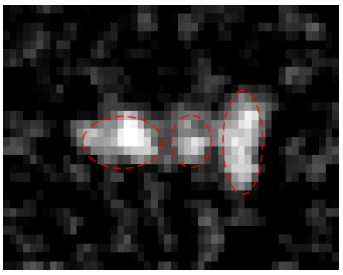
Firstly, enlarged image details need further inspection in order to evaluate image quality. Figure 2.9 shows ROI (region of interest) results reconstructed by three different experiments. There are three separable ellipses of low contrast in the small box area, as shown in Figure 2.9a and 2.9b. In all the three experiments, α values are chosen manually from the indicated range in Table 2.4 to achieve the best ROI image (with similar overall image quality). Comparing Figure2.9c, 2.9d and 2.9e, it can be seen that the three elliptic regions become inseparable in all the three experiments, however in the two TV model experiments, attenuation coefficients in the low contrast regions are reconstructed more uniformly. It could also be seen that in the 290-projection-TV experiment (Figure2.9d) the outline of the low contrast regions is slightly better reconstructed compared to the 58-projection-TV experiment (Figure2.9e). Therefore, Figure 2.9 demonstrates that complicated low contrast details could be better reconstructed by using the TV-regularized model, and that ROI results could be slightly improved by increasing the number of projection measurements.



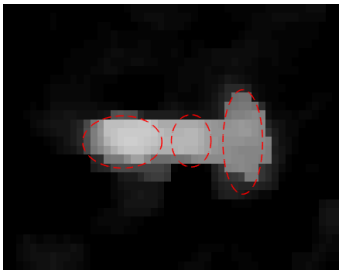
(a) The true full image. Red box indicates the ROI region.



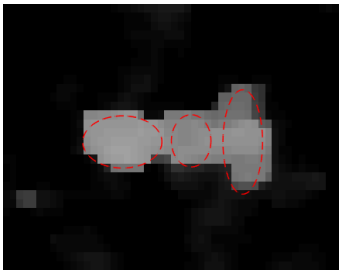
(b) The true ROI image.



(c) Tikhonov model, 290 projections, $\alpha = 3000^2$



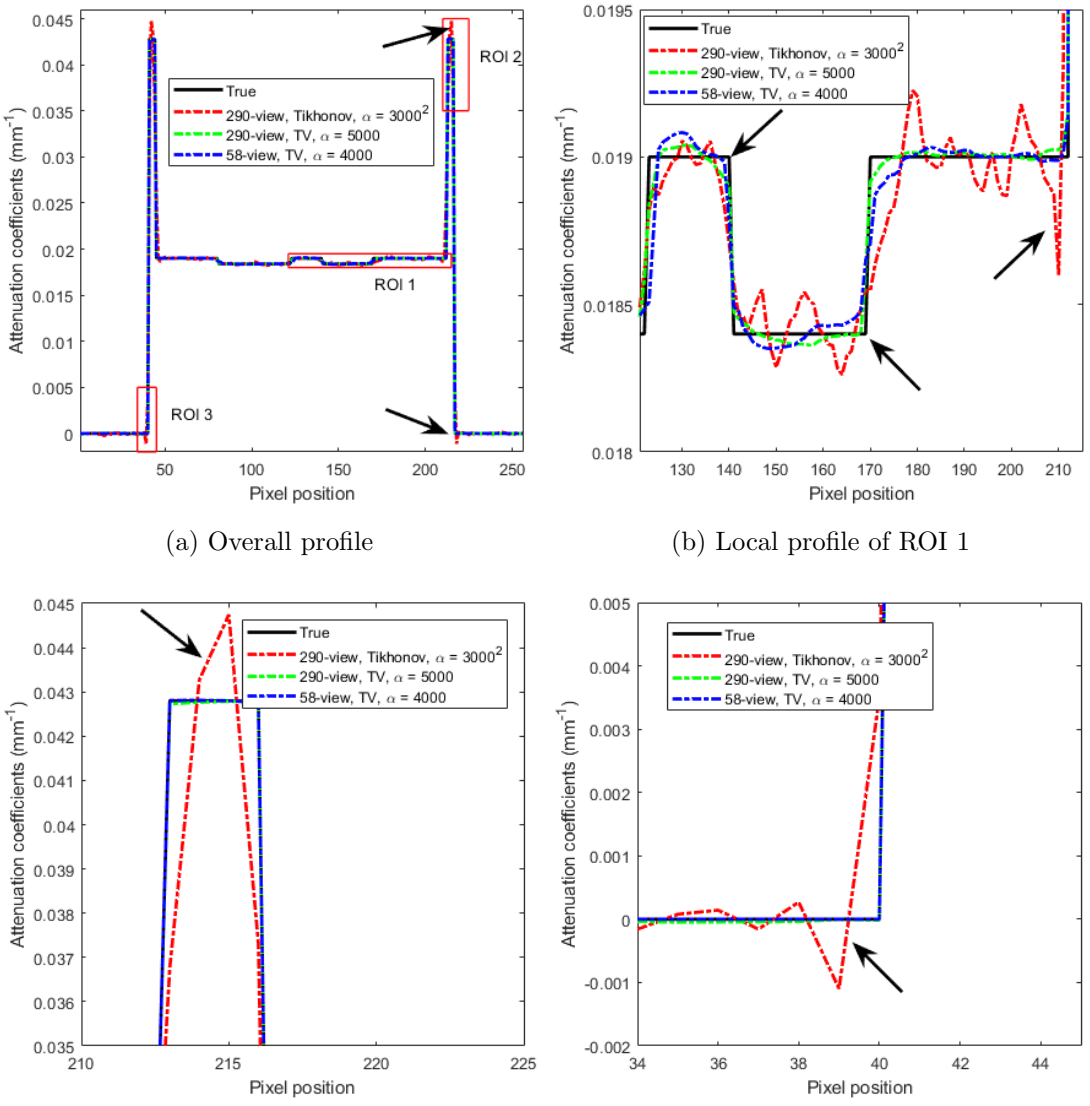
(d) TV model, 290 projections, $\alpha = 5000$



(e) TV model, 58 projections, $\alpha = 4000$

Figure 2.9: ROIs of the images reconstructed by 3 different experiments. Red ellipses indicate low contrast (blood) regions. Display window $[0.0190, 0.0195]\text{mm}^{-1}$.

Secondly, one-dimensional profile plots can be used to further visualize the difference between experiments and especially difference at edge locations. In Figure 2.10, horizontal profiles of the reconstructed images are plotted across the 133th row, and the profile from the true image is also given for reference. Three ROIs are selected to further visualize the difference at edge locations as shown in Figure2.10b, 2.10c and 2.10d. The profile results show that the TV-regularized model preserves both strong (brain-bone, bone-air as shown in Figure2.10c and 2.10d) and weak edges (water-brain as shown in Figure2.10b) much better than Tikhonov-regularized model. It can also be observed in Figure2.10b that increasing the projection number slightly improves the weak edge preservation for TV-regularized experiments.



(a) Overall profile

(b) Local profile of ROI 1

(c) Local profile of ROI 2

(d) Local profile of ROI 3

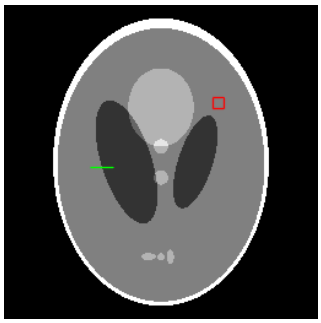
Figure 2.10: Horizontal profile (row 133) of images reconstructed by different experiments. Local profiles of the three ROIs indicated in (a) are plotted in (b), (c) and (d).

Finally, a resolution and noise trade-off study is conducted in order to quantitatively evaluate the image quality in different experiments. In quantitative evaluations, the two parameters, resolution and noise, need to be monitored at the same time, and an image with better quality should have both higher resolution and lower noise level. In regularized iterative reconstructions, the regularization parameter α plays a significant role in balancing image resolution and

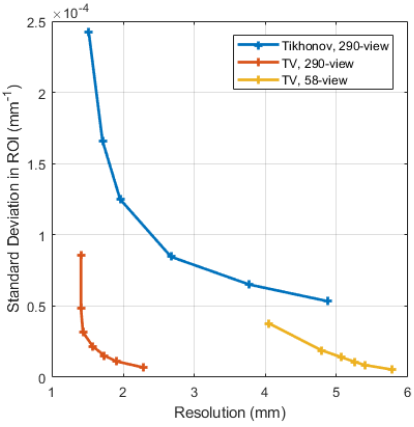
noise. Therefore, resolution-noise trade off plots are required for comparing results in different experiments.

Given a reconstructed image, the image resolution can be calculated by matching the edge spread function (ESF) to the reconstructed result. This calculation procedure is explained in [30] and [31]. A local profile passing through a low contrast region is firstly selected, as the green line shown in Figure 2.11a. An edge broadening kernel is then selected as a Gaussian function with standard deviation σ_b, and Gaussian filtered profile can be computed by a convolution between the true profile and the Gaussian kernel. The edge spread function has the form of measuring the error between the Gaussian filtered profile (red and purple curve in Figure 2.11c) and the reconstructed profile (blue and orange curve in Figure 2.11c) parameterized by σ_b. An optimal σ_b is obtained by minimizing ESF, and the full-width at half-maximum (FWHM) of the Gaussian broarding kernel, which is $2.35\sigma_b$, is the measured resolution. Furthermore, the image noise can be obtained by computing the standard deviation of a small uniform ROI, as the red box shown in Figure 2.11a.

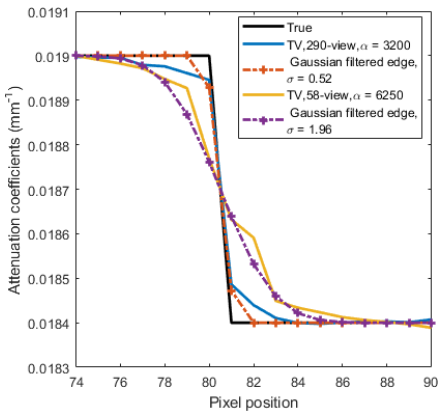
A resolution-noise pair of a reconstructed image can be computed as above, and a resolution-noise tradeoff curve can be obtained by computing resolution-noise pairs using various regularization parameter α values. Figure 2.11b shows the resolution-noise curves for 3 different experiments. It can be seen that the TV-regularized result outperforms the Tikhonov-regularized result in terms of resolution and noise. It can also be seen that increasing the projection number greatly improves the image resolution for TV-regularized experiments by comparing the red and orange curve.



(a) Illustration of local uniform ROI (red) and step edge profile (green) for noise and resolution computation.



(b) Resolution - noise tradeoff curves parameterized by regularization parameter for 3 different experiments.



(c) Guassian filtered profile plots vs reconstructed profile plots.

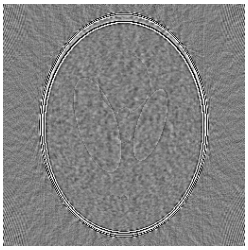
Figure 2.11: Resolution - noise tradeoff plot.

Convergence analysis

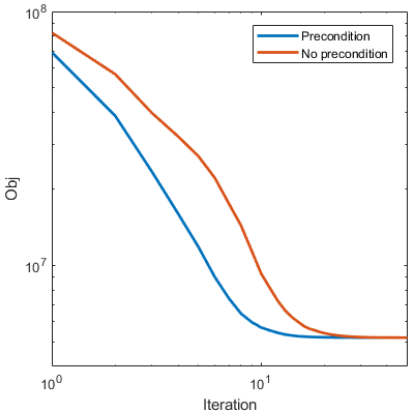
In this subsection, acceleration techniques and the convergence analysis are discussed for both Tikhonov and TV-regularized experiments. For experiments without further specification, a Hann-filtered FBP result is used for initial guess. Without loss of generality, only results in 290-projection experiments are used here, and α is set to be 2000^2 for the Tikhonov experiment and 4000 for the TV experiment. 58-projection experiments shows similar convergence properties and thus are skipped.

Tikhonov model For the Tikhonov-regularized model (2.22), quasi-newton method with steepest descent is applied for minimizing the objective function. In order to further accer-

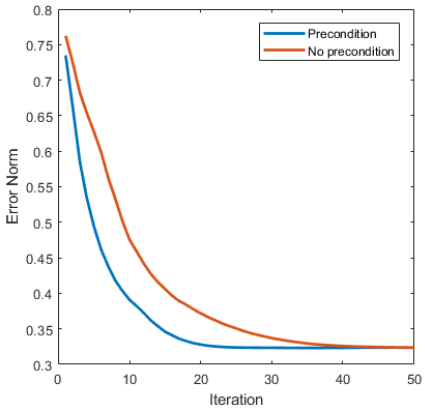
late the convergence, a diagonal preconditioning matrix $\boldsymbol{P}$ defined as (2.29) is applied and the preconditioned model (2.28) is then solved using the same numerical method. Figure 2.12 summaries the convergence analysis results for Tikhonov experiment. Figure 2.12a shows the difference between the reconstructed image after 50 iterations u^* and the true image u_{True}, and it shows that the Tikhonov model is incapable of preserving strong edges. Figure 2.12b shows the objective function curves with and without preconditioning. It shows that the objective function converges faster in earlier iterations when applying preconditioning, and that the objective function converges to the same result with or without precondition. In order to further ensure convergence towards the true image, the error norm, $\|u^* - u_{True}\|_2$, is computed after every iteration, and error norm curves are plotted in Figure 2.12c for precondition and no-precondition experiment. It can be observed that the error norm decreases during iterations for both precondition and non-precondition experiments, and that the error norm converges also faster when applying preconditioning.



(a) $u^* - u_{True}$, display window [-0.005 0.005] mm^{-1}



(b) Obj vs Iteration (log-log)



(c) $\|u^* - u_{True}\|_2$ vs Iteration

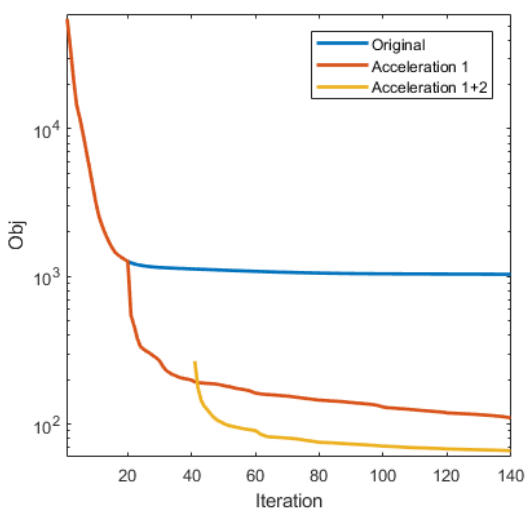
Figure 2.12: Convergence analysis for Tikhonov experiment.

TV model For TV-regularized model (2.31), two numerical methods, ADMM (Algorithm 4) for solving splitted TV model (2.32) and quasi-Newton (Algorithm 3) with backtracking line search for solving smoothed TV model (2.37), are firstly tested in this subsection. It is shown that after sufficient iterations, the two methods converge to the same result as long as the smoothing parameter ϵ for (2.37) is small enough. For all experiments in this chapter, ϵ is chosen to be 10^{-10}. For the ADMM method, the augmented Lagrangian parameter ρ influences the convergence rate significantly [21]. In the following, convergence of the quasi-Newton method for smoothed TV model is studied similar as the convergence for Tikhonov model. Several acceleration techniques are briefly explained and applied here, including:

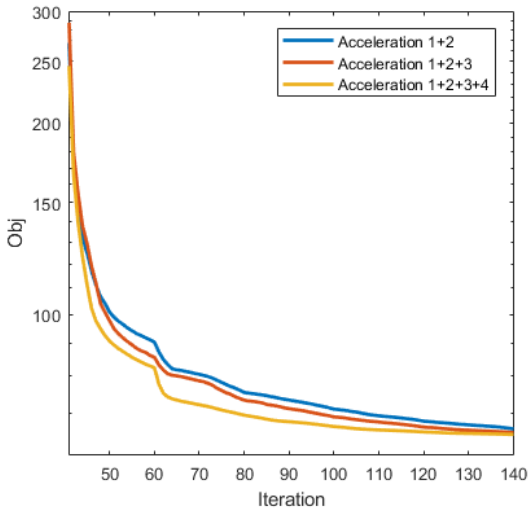
- Acceleration 1: periodic restart quasi-Newton;
- Acceleration 2: unweighted iterative result as a new initial guess;
- Acceleration 3: circular reconstruction region;
- Acceleration 4: preconditioning.

Firstly, as illustrated in theoretical section 2.4, the convergence of a quasi-Newton algorithm becomes very slow when the search direction and the gradient descent direction are nearly perpendicular to each other. Therefore, a restart of the quasi-Newton algorithm (Acceleration 1) is required to maintain a proper search direction during each iteration. The effectiveness

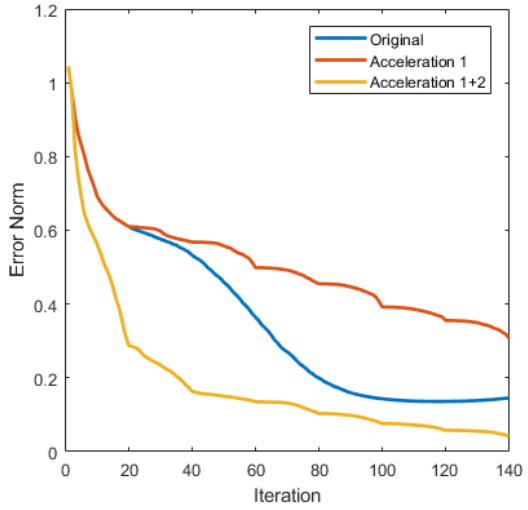
of Acceleration 1 method is proven by the blue and red curve in Figure 2.13a. It can be observed that convergence rate starts to accelerate again after periodic restart of the quasi-Newton algorithm. In this experiment, the algorithm is restarted every $N_{restart} = 20$ iterations. Unlike the increased convergence of objective function, it can be observed in Figure 2.13c that the decrease of error norm becomes much slower at early iterations after only applying the periodic restart. Only when more acceleration techniques are applied, the convergence of the error norm becomes faster than the original.



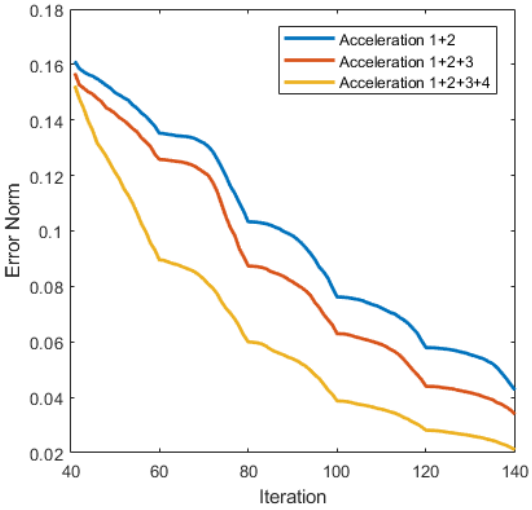
(a) Obj vs Iteration.



(b) Obj vs Iteration.



(c) Error norm vs Iteration.



(d) Error norm vs Iteration.

Figure 2.13: Convergence analysis for TV experiment.

Secondly, a new initial guess other than an FBP result can be applied to further accelerate the convergence for the approximated-TV problem. One reason causing the slow convergence of the weighted and regularized least-squares problem (2.31) is the large condition number of the covariance matrix Σ. Therefore, one way to achieve faster convergence is to firstly solve an unweighted problem by setting Σ to be an identity matrix (a proper α needs to be re-selected here), and then use this result as an initial guess to solve the exact weighted problem (2.31). Images reconstructed by solving the unweighted problem with $\Sigma = I$ have poorer image quality compared to images reconstructed from exact weighted problem (2.31), however roughly solving

the unweighted problem requires much less iterations. In the experiment, 40 iterations of quasi-Newton with $N_{restart} = 20$ are firstly run to solve the unweighted problem, and the result is used for an initial guess to solve the exact problem with another 100 iterations of quasi-Newton. Note that the objective function plot of this method starts from the 41th iteration, because the first 40 iterations are for obtaining the initial guess by optimizing a different objective function. The effectiveness of this Acceleration 2 method is proven by the red and orange curve in Figure 2.13a and 2.13c. Both objective function and error norm converge much faster after applying a new initial guess.

Thirdly, based on the geometry setting in subsection 2.5.1, not all the $256 * 256$ pixels in the 300×300 mm^2 square region can be reconstructed by FBP, because a pixel can be reconstructed only when it is measured in all different angular projections. Therefore in the current geometric setting, only pixels inside a circular region with a radius of 150mm can be reconstructed, and all other unrealistic pixels can be excluded from iterative reconstruction experiments, assuming they all have values of 0. By removing unnecessary pixels, not only the amount of calculation per iteration is reduced (system matrix $\boldsymbol{W}$ reduces to 78% of the original), but the convergence rate versus iteration is sped up as well. The effectiveness of this Acceleration 3 method is proven by the blue and red curve in Figure 2.13b and 2.13d. Both objective function and error norm converge faster after removing unnecessary pixels from reconstruction.

Finally, a preconditioning technique can also be used for accelerating the approximated TV-model (2.37) problem. For the Tikhonov-regularized model (2.22), a preconditioning matrix $\boldsymbol{P}$ can be defined by equation (2.29), and has been proven very useful for convergence acceleration. Similarly, a preconditioning matrix $\boldsymbol{P}$ for approximated TV-model (2.37) is defined as

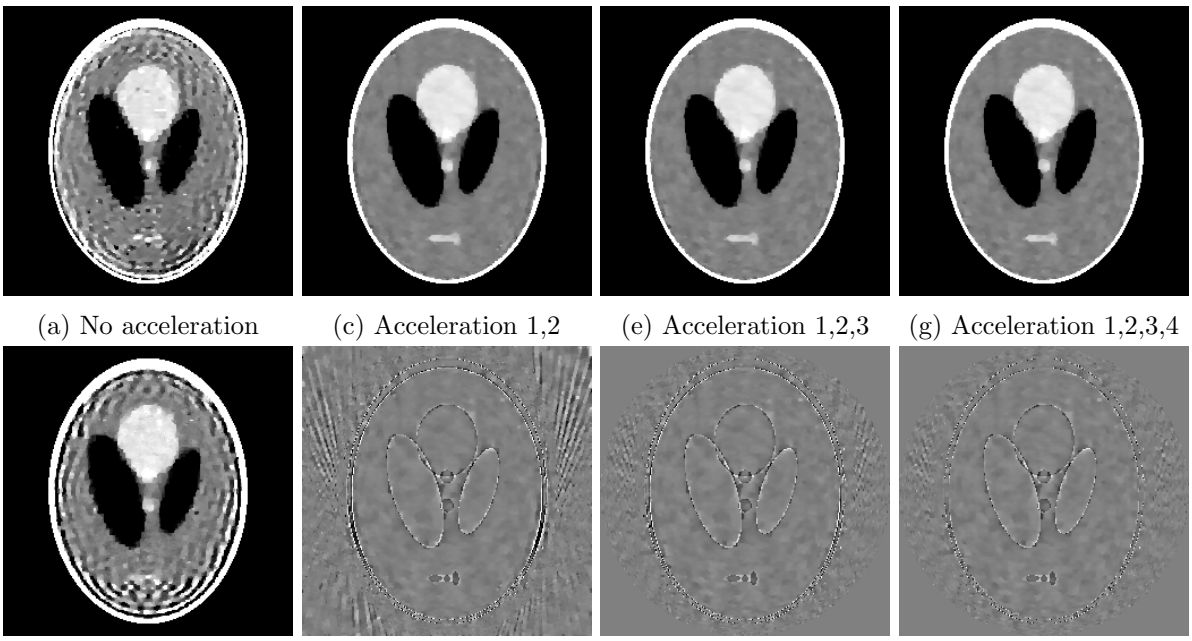
$$\boldsymbol{P} = \operatorname{diag}\left\{\boldsymbol{W}_{\Sigma}^{T}\boldsymbol{W}_{\Sigma}\mathbf{1} + 2\alpha\nabla^{2}\widetilde{J}_{ani}(\boldsymbol{u}^{(0)})\mathbf{1}\right\}^{-1/2}. \tag{2.38}$$

Note that $\boldsymbol{P}$ defined above depends on the initial guess of image, $\boldsymbol{u}^{(0)}$, therefore it can be updated every time the quasi-Newton algorithm restarts. The effectiveness of Acceleration 4 method is proven by the red and orange curve in Figure 2.13b and 2.13d.

In summary, all of the four acceleration methods are capable of speeding up the quasi-Newton algorithm for solving the approximated TV problem, and images with good quality could be achieved by applying a combination of them. Figure 2.14 shows reconstruction results after 140 iterations, and it can be seen that images with better quality can be reconstructed after more acceleration methods are applied. To better visualize the difference between result 2.14c, 2.14e and 2.14g, difference images between reconstruction results and the true image are plotted in 2.14d, 2.14f and 2.14h. It can be seen from the difference images that pixels near edges are better reconstructed when more acceleration methods are applied. Reconstruction results after 80 iterations are also shown in Figure 2.15, when none of the experiments, Acceleration 1+2, Acceleration 1+2+3 and Acceleration 1+2+3+4, is nearly converged. This suggests if all of the 4 acceleration techniques are applied, an image with reasonable quality can be reconstructed with an early termination of the algorithm, which reduces the running time dramatically. However, how to choose a proper stopping criterion remains a problem.

2.5.3 Discussion and conclusion

In Section 2.5, statistical reconstruction models with two different regularization terms, Tikhonov or anisotropic TV, are tested using simulated noisy data. It is shown that Tikhonov-regularized model can only be used for sinogram data with sufficient projections, whereas the proposed TV model works well for both sufficient-projection or few-projection data. To achieve an image with required quality, regularization parameter should be selected carefully, and L-curve serves as an efficient method for both Tikhonov and TV model problems. In order to better evaluate image quality, several commonly used techniques are introduced and applied to experimental



(a) No acceleration (c) Acceleration 1,2 (e) Acceleration 1,2,3 (g) Acceleration 1,2,3,4

(b) Acceleration 1 (d) Acceleration 1,2 (f) Acceleration 1,2,3,4 (h) Acceleration 1,2,3,4

Figure 2.14: Results after 140 iterations for approximated TV reconstruction. Display window of (a), (b), (c), (e) and (g) is [0.0185 0.0195] mm^{-1}. Figure (d), (f) and (h) show the difference between reconstruction and true image, and display window is [-0.0005 0.0005]mm^{-1}.

(a) Acceleration 1+2 (b) Acceleration 1+2+3 (c) Acceleration 1+2+3+4

Figure 2.15: Results after 80 iterations for approximated TV reconstruction. The algorithm is terminated early before convergence. Display window [0.0185 0.0195]mm^{-1}.

results for both models, including ROI plots, one-dimensional profile plots and resolution-noise tradeoff curve. Qualitative and quantitative evaluation of the reconstruction results shows that images can be slightly better reconstructed with more projection data using TV model when the total incident energy remains the same. However, one disadvantage of using more projection data is the longer running time caused by a larger dataset size, so a balance between image quality and algorithm speed should be considered when designing CT scanning protocols.

In this chapter, a precondition technique called matrix scaling is used to accelerate reconstruction for both Tikhonov- and TV- regularized problems, and the efficiency for this preconditioning method has been proved by numerically experiments. For any least-squares problem (2.25), let the precondition matrix $\boldsymbol{P} = \text{diag}\{\widetilde{\boldsymbol{W}}\boldsymbol{1}\}^{-1/2}$. The Hessian matrix for the original problem is $\widetilde{\boldsymbol{W}}^T\widetilde{\boldsymbol{W}}$, and for the preconditioned problem (2.28) will be $\boldsymbol{P}\widetilde{\boldsymbol{W}}^T\widetilde{\boldsymbol{W}}\boldsymbol{P}$. It can be easily proved that matrix $\boldsymbol{P}\widetilde{\boldsymbol{W}}^T\widetilde{\boldsymbol{W}}\boldsymbol{P}$ and $\widetilde{\boldsymbol{W}}^T\widetilde{\boldsymbol{W}}\boldsymbol{P}^2$ have the same spectrum, and that columns of $\widetilde{\boldsymbol{W}}^T\widetilde{\boldsymbol{W}}\boldsymbol{P}^2$ all have the same L^1-norms. A lower bound for the condition number of a matrix is given by the following theorem.

Theorem 2 *(Poor scaling of a matrix)*
Given any full rank matrix $\boldsymbol{A}$ of size $n \times n$, let $\boldsymbol{a}_1, \boldsymbol{a}_2, ..., \boldsymbol{a}_n$ be the n column vectors of $\boldsymbol{A}$, and the condition number of $\boldsymbol{A}$, $C(\boldsymbol{A})$, is given by the ratio of the maximum and minimum eigenvalue, or equivalently $\|\boldsymbol{A}\|_2 \, \|\boldsymbol{A}^{-1}\|_2$. Then the lower bound for $C(\boldsymbol{A})$ is given by

$$C(\boldsymbol{A}) \geq \frac{\|\boldsymbol{a}_i\|_2}{\|\boldsymbol{a}_j\|_2}, \forall i, j \in [1, n]. \tag{2.39}$$

The above theorem suggests a possibly better preconditioning matrix could be designed to balance the L^2-norms of the column vectors instead of L^1-norms. The relationship between $C(\widetilde{\boldsymbol{W}}^T\widetilde{\boldsymbol{W}}\boldsymbol{P}^2)$ and $C(\widetilde{\boldsymbol{W}}^T\widetilde{\boldsymbol{W}})$ has been discussed and proved in some early literature [32, 33]. Usually the condition number becomes smaller after precondition when $\widetilde{\boldsymbol{W}}$ is poorly scaled, however this is not always true for any matrix $\widetilde{\boldsymbol{W}}$.

Quasi-Newton method has been proved to have superlinear convergence [5]. A recent literature [34] has shown that different types of quasi-Newton methods all belong to the first-order methods, suggesting that the worst-case convergence rate for quasi-Newton methods is linear and that the local superlinear convergence cannot be achieved until the number of iterations reaches half the size of the problem dimension. Even though the convergence rate for the exact Newton method will be quadratic and not depend on condition number, the convergence rate for a quasi-Newton method will still depend on the condition number of the Hessian, and thus precondition technique which reduces the condition number will be effective for accelerating quasi-Newton methods.

One other thing observed during experiments is that approximated TV-regularized problem requires many more iterations to converge when using the quasi-Newton algorithm described in Algorithm 3, compared to Tikhonov-regularized problem. Even though approximated TV model has been used for image reconstruction problems in several references[22, 23], research on how to efficiently solve the smoothed TV problem has been rare as far as I know. Results in this chapter show that by using quasi-Newton algorithm with a combination of four acceleration techniques, statistical CT reconstruction problem with TV term can be roughly solved within 100 iterations. This result is novel and very promising comparing to existing results in reference [22]. However, further experiments are required to compare the performance of this accelerated quasi-Newton algorithm with primal-dual algorithms for solving exact TV model problems.

Besides the low detector sampling rate and the noisy sinogram data, there may be one other reason causing the poor resolution in the ROI images Figure 2.9d and 2.9e. According to [35] (see Fig. 4 on Page 11), anisotropic TV term used in this thesis favours corners and rectangular structures, which may not be a good choice for reconstructing the small ellipses in the image.

Image quality may be improved by using isotropic TV regularization which favours rounded edges, or even further improved by using a linear combination of anisotropic and isotropic TV terms [17].

Chapter 3

Multi-energy reconstruction - A two-step algorithm

This chapter starts to focus on multi-energy reconstruction problems. Unlike the single-energy problem in which the data is measured at one given energy E, multi-energy CT measures projection data in multiple energy bins B_1, B_2,..., B_{N_e}. Adding the energy dimension not only increases the dataset size, but also allows us to discover more quantitative metrics by exploring the relationship between data at different energies. This chapter will start from introducing the physical basics for multi-energy CT, including the material decomposition and the photon detection. A general framework for spectral reconstruction problems is then formulated, and spectral reconstruction models are categorized into two types: two-step models and one-step models. Finally an algorithm for solving a two-step model is developed. Numerical experiments using simulated noisy data show that the material decomposition problem (step one) can be most accurately solved by combining a maximum likelihood model with an approximated weighted least-squares model, and experiments also show that the basis image reconstruction problem (step two) can be solved efficiently by using numerical algorithms developed in Chapter 2.

3.1 Multi-energy X-ray attenuation

3.1.1 Decomposition of attenuation coefficients

Let $u(E, \vec{z})$ be the attenuation coefficient at location $\vec{z}$ and energy E, and it can be expressed as a linear combination of basis functions $f_i(E), i = 1, 2, ..., N_b$, as given by

$$u(E, \vec{z}) = \sum_{i=1}^{N_b} A_i(\vec{z}) f_i(E). \tag{3.1}$$

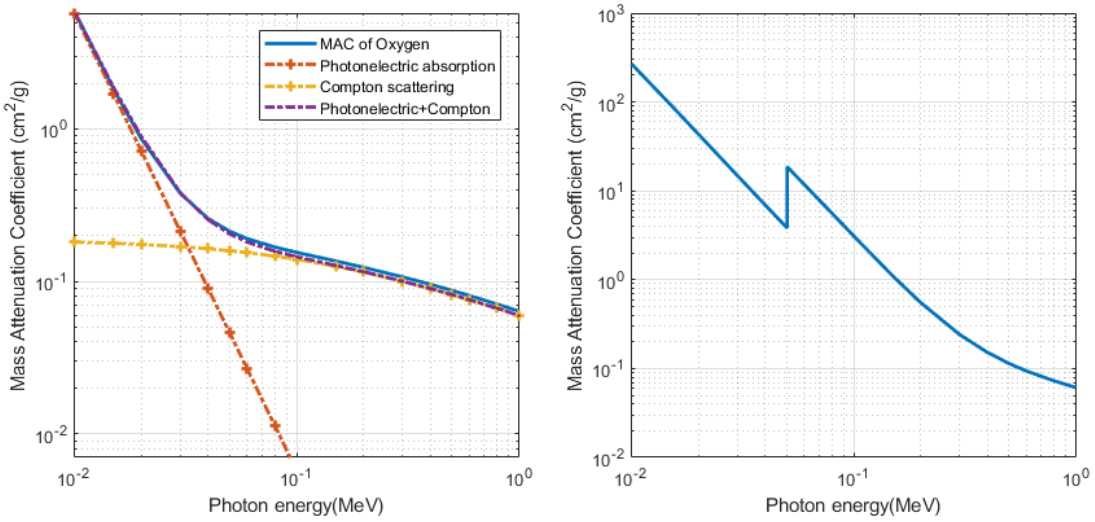
Here $f_i(E)$ is the ith element for composing $u(E, \vec{z})$ at energy E, $A_i(\vec{z})$ independent of energy is the coefficient for $f_i(E)$, and N_b is the total number of basis for decomposition. Without the existence of K-edge elements (see below), $u(E, \vec{z})$ can be expressed as a combination of two basis functions, $f_1(E) = f_{\text{ph}}(E)$ and $f_2(E) = f_{\text{KN}}(E/E_e)$, which are given by [36]

$$f_{\text{ph}}(E) = \frac{1}{E^3}, \tag{3.2}$$

and

$$f_{\text{KN}}(\alpha) = \frac{1+\alpha}{\alpha^2}\left[\frac{2(1+\alpha)}{1+2\alpha} - \frac{1}{\alpha}\ln(1+\alpha)\right] + \frac{1}{2\alpha}\ln(1+2\alpha) - \frac{1+3\alpha}{(1+2\alpha)^2}, \alpha = \frac{E}{E_e} = \frac{E}{510.975\text{keV}}. \tag{3.3}$$

Here, $A_1(\vec{z})f_{\text{ph}}(E)$ is the photoelectric absorption component, and $A_2(\vec{z})f_{\text{KN}}(E/E_e)$ is the Compton scattering component for attenuation coefficient. As an example, mass attenuation coefficient (MAC) of oxygen (O) element can be decomposed into two components, as shown in Figure 3.1a. The MAC curve for oxygen and all other MAC data used in this chapter are read from XCOM online database [37].



(a) MAC of Oxygen(O) and decomposition result.

(b) MAC of Gadolinium (Gd).

Figure 3.1: Mass attenuation coefficient (MAC) for Oxygen and Gadolinium. In (a), oxygen MAC can be decomposed into two components, photonelectric absorption (red) and Compton (orange) scattering. In (b), a sharp edge can be observed around 50keV energy.

The K-edge discontinuity is the sharp edge in the absorption spectrum of an element, and it occurs when the incident photon energy equals to the binding energy of the K-shell electron of an atom (See more in Section 2.3.2 in [38]). A K-edge element is an element with K-edge discontinuity within the X-ray energy range of interest. For example, the MAC of the Gadolinium (Gd) element is plotted in Figure 3.1b, and K-edge energy of Gadolinium is at 50.2 keV. With existence of a K-edge element, the two basis defined above, $f_{\text{ph}}(E)$ and $f_{\text{KN}}(\alpha)$, are not enough to model the attenuation coefficient $u(E,\vec{z})$ because of the K-edge discontinuity. Therefore, a third basis $f_3(E)$, which is the mass attenuation coefficient of the K-edge element, is added for the decomposition, and $A_i(\vec{z})$ will be local mass density for the specific element. More basis will be required if more than one K-edge elements exist.

Similar to (2.1), the projected attenuation at location (θ,ξ) in the polar coordinate system, $p_k(E,\theta,\xi)$, is a line integral of $u(E,\vec{x})$, expressed as $p(E,\theta,\xi)=\int_{L(\theta,\xi)}u(E,\vec{x})ds$. Substitute $u(E,\vec{x})$ by equation (3.1) and $p(E,\theta,\xi)$ can be further expressed as

$$p(E,\theta,\xi)=\int_{L(\theta,\xi)}\sum_{i=1}^{N_b}A_i(\vec{z})f_i(E)ds=\sum_{i=1}^{N_b}\left(f_i(E)\int_{L(\theta,\xi)}A_i(\vec{z})ds\right)\triangleq\sum_{i=1}^{N_b}f_i(E)\cdot Y_i(\theta,\xi). \quad (3.4)$$

Here $Y_i(\theta,\xi)=\int_{L(\theta,\xi)}A_i(\vec{z})ds$ denotes the projected decomposed coefficient for the ith basis f_i at polar system location (θ,ξ). The above equation (3.4) indicates a linear relationship between the projected attenuation and decomposition coefficients, showing that the decomposition can be conducted in the image domain, as well as the projection domain. Relationship between attenuation coefficient in space $u(E,\vec{z})$, decomposition coefficient in space $A_i(\vec{z})$, projected attenuation $p(E,\theta,\xi)$ and projected decomposition coefficient $Y_i(\theta,\xi)$ is summarized in Figure 3.2.

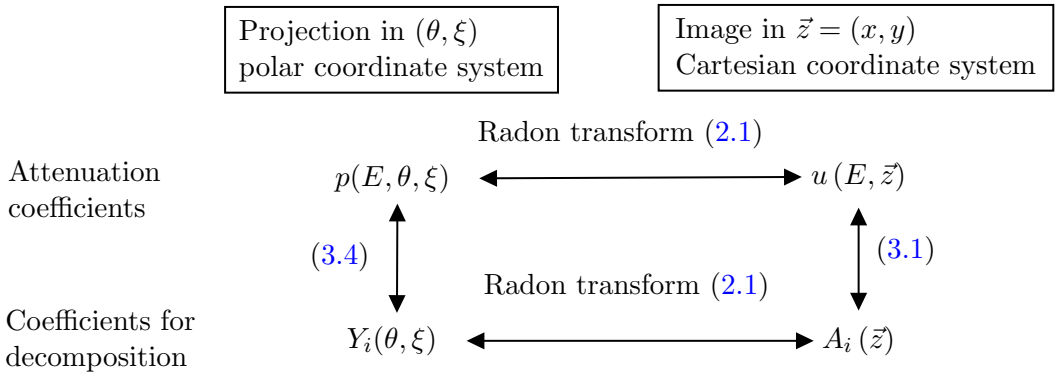


Figure 3.2: Illustration of attenuation decomposition in projection domain and in image domain.$i = 1, 2, ..., N_b$.

3.1.2 Photon detection and energy bins

In previous Chapter 2, equation (2.16) shows an exponential relationship between the expected photon number and projected attenuation at given energy E. For multi-energy cases, let $\bar{I}(B, \theta, \xi)$ be the expected value of the number of photons in energy bin B at projection location (θ, ξ), and it is an integral over energy, expanded as below

$$\begin{aligned} \bar{I}(B, \theta, \xi) &= \int_{E \in B} \bar{I}(E, \theta, \xi) D(E) dE = \int_{E \in B} I_{in}(E) e^{-p(E, \theta, \xi)} D(E) dE \\ &\approx e^{-p(B, \theta, \xi)} \int_{E \in B} I_{in}(E) D(E) dE. \end{aligned} \tag{3.5}$$

Here $I_{in}(E)$ denotes the x-ray incident photon number at energy E, and $D(E)$ denotes the detector absorption efficiency. Let $p(B, \theta, \xi)$ be the averaged $p(E, \theta, \xi)$ over energy bin B, and the approximate equality holds only when $p(E, \theta, \xi)$ varies very little in B. When the approximation is true, $p(B, \theta, \xi)$ can be computed by

$$p(B, \theta, \xi) \approx \ln \left(\frac{\int_{E \in B} I_{in}(E) D(E) dE}{\bar{I}(B, \theta, \xi)} \right) \triangleq \ln \left(\frac{I_{in}(B)}{\bar{I}_k(B, \theta, \xi)} \right). \tag{3.6}$$

Unlike the nonlinear relationship between photon number and projected attenuation, the linear decomposition relationship (3.1) and (3.4) still hold after an integration of E in energy bin B, expressed as

$$u(B, \vec{z}) \triangleq \frac{1}{|B|} \int_{E \in B} u(E, \vec{z}) \, dE = \sum_{i=1}^{N_b} A_i(\vec{z}) \left(\frac{1}{|B|} \int_{E \in B} f_i(E) dE \right), \tag{3.7}$$

$$p(B, \theta, \xi) \triangleq \frac{1}{|B|} \int_{E \in B} p(E, \theta, \xi) dE = \sum_{i=1}^{N_b} Y_i(\theta, \xi) \left(\frac{1}{|B|} \int_{E \in B} f_i(E) dE \right), \tag{3.8}$$

where $|B|$ denotes the width of energy bin B.

In a real scanning, the detected photon number $I(B, \theta, \xi)$ in energy bin B at location (θ, ξ) also follows a Poisson distribution $I(B, \theta, \xi) \sim Poisson(\bar{I}(B, \theta, \xi))$. Replacing $\bar{I}(B, \theta, \xi)$ by measured data $I(B, \theta, \xi)$ in equation (3.6), projected attenuation $p(B, \theta, \xi)$ is obtained, which can be used for image reconstruction as in Chapter 2. For monochromatic CT reconstruction, equation (3.6) is commonly used to obtain projected attenuation, the same as step 4 in Algorithm 5 in the previous chapter, and the inaccuracy of the approximation may cause the beam hardening effect when the energy bin is not narrow enough. For polychromatic CT, according

to their energy, output x-ray photons are collected in one of the N_e energy bins $B_1, B_2, ..., B_{N_e}$, and N_e sets of sinogram data within different energy bins, $p(B_1, \theta, \xi)$, $p(B_2, \theta, \xi)$,..., $p(B_{N_e}, \theta, \xi)$, are measured and N_e attenuation images can be reconstructed separately from sinogram data of different energies.

3.2 Multi-energy reconstruction

3.2.1 Why multi-energy reconstruction is needed

Reconstruction algorithms discussed in the previous chapter can be directly used for reconstructing attenuation images at different energy bins, as briefly discussed at the end of last section. However, there is still a need to further develop algorithms specifically for multi-energy CT reconstruction, at least for the following reasons. Firstly, computing projected attenuation $p(B, \theta, \xi)$ from detected photon numbers by equation (3.6) may not be accurate. For single-energy CT scanning , it is possible to choose one energy bin within which attenuation $p(E, \theta, \xi)$ varies very little for different energy E. However for multi-energy scanning, it is difficult to select more than one energy bins and ensure that attenuation $p(E, \theta, \xi)$ varies very little for different energy E within every bin, unless all energy bins are narrow enough. Therefore, developing algorithms to use measured photon numbers $I(B, \theta, \xi)$ rather than projections $p(B, \theta, \xi)$ for reconstruction may be more accurate.

Secondly, attenuation sinograms at different energy bins are correlated as they are measurements for the same object, and the reconstructed attenuation images at different energy bins are also correlated. Such inter-energy correlation can be modeled by the attenuation decomposition relationship into a multi-energy reconstruction problem in order to improve the image quality [39]. Furthermore, when total incident photon energy remains fixed and more energy bins are selected for measuring output x-ray photons, less photons will be collected in one single energy bin, leading to noisier measurements. Noisier measurements for multi-energy CT also suggests the necessity of modeling the inter-energy correlation for noise suppression.

Finally, in previous section 3.1.1, it is shown that there exists a linear relationship between attenuation coefficients $u(E, \vec{z})$ and decomposition coefficients $A_i(\vec{z})$, and $A_i(\vec{z})$ can be local mass density for a K-edge element given the MAC curve as a decomposition basis. This suggests that using multi-energy data, not only attenuation coefficient images at different energy can be reconstructed, but the mass density image of a K-edge element can be reconstructed as well. Therefore, algorithms need to be developed for reconstructing the mass density images, which could provide more information about the scanning object.

3.2.2 Problem formulation

In this subsection, frameworks for different types of multi-energy reconstruction algorithms are formulated in a discrete setting, and several existing algorithms in reference are briefly introduced as examples for different types of algorithms.

Firstly, variables used in multi-energy reconstruction are p, I, Y, u and A, as explained in section 3.1 and especially in Figure 3.2. Discretization of the five variables is given as below

$$\begin{aligned}
\boldsymbol{P} &= [p_{j,k}]_{N_p \times N_e}, p_{j,k} \triangleq p(B_k, \theta_j, \xi_j); \\
\boldsymbol{I} &= [I_{j,k}]_{N_p \times N_e}, I_{j,k} \triangleq I(B_k, \theta_j, \xi_j); \\
\boldsymbol{Y} &= [Y_{j,i}]_{N_p \times N_b}, Y_{j,i} \triangleq Y_i(\theta_j, \xi_j); \\
\boldsymbol{U} &= [u_{j,k}]_{N_i \times N_e}, u_{j,k} \triangleq u\left(B_k, \vec{z}_j\right); \\
\boldsymbol{A} &= [A_{j,i}]_{N_i \times N_b}, A_{j,i} \triangleq A_i\left(\vec{z}_j\right);
\end{aligned} \tag{3.9}$$

where N_p is the number of pixel in the projection domain, N_i is the number of pixel in the image domain, N_e is the number of energy bins and N_b is the number of basis used for attenuation

decomposition. Note that columns of matrix $\boldsymbol{P}$, $\boldsymbol{I}$, and $\boldsymbol{U}$ are data at different energies, and columns of matrix $\boldsymbol{Y}$ and $\boldsymbol{A}$ are data for different decomposition basis.

System matrix $\boldsymbol{W}$ for CT geometry has been introduced in section 2.2, and the size is $N_p \times N_i$ here. Therefore, the radon transform relationship between $\boldsymbol{P}$ and $\boldsymbol{U}$, $\boldsymbol{Y}$ and $\boldsymbol{A}$ can be expressed as

$$\boldsymbol{P} = \boldsymbol{W}\boldsymbol{U}, \boldsymbol{Y} = \boldsymbol{W}\boldsymbol{A}. \tag{3.10}$$

Let $\boldsymbol{F}$ be the decomposition basis matrix of size $N_b \times N_e$ defined as below

$$\boldsymbol{F} = [f_{i,j}]_{N_b \times N_e}, f_{i,j} \triangleq f_i(B_j) \triangleq \frac{1}{|B_j|} \int_{E \in B_j} f_i(E) dE, \tag{3.11}$$

and the decomposition relationship (3.7) and (3.8) can be summarized in matrix form as

$$\boldsymbol{P} = \boldsymbol{Y}\boldsymbol{F}, \boldsymbol{U} = \boldsymbol{A}\boldsymbol{F}. \tag{3.12}$$

The multi-energy reconstruction problem now can be described as finding the optimal decomposition images as columns of matrix $\boldsymbol{A}$, given system matrix $\boldsymbol{W}$, decomposition basis matrix $\boldsymbol{F}$ and data matrix $\boldsymbol{I}$. It can be generalized as an optimization problem as below

$$\boldsymbol{A}^* = \arg\min_{\boldsymbol{A}} J\left(\boldsymbol{A}; \boldsymbol{W}, \boldsymbol{F}, \boldsymbol{I}\right), \tag{3.13}$$

where $J : \mathbb{R}^{N_i \times N_b} \to \mathbb{R}$ is the objective function, and it usually contains a data fidelity term and a regularization term, similar as problems discussed in Section 2.4. Note that sometimes photon measurement matrix $\boldsymbol{I}$ is not directly used in problem (3.13), but it will be firstly transformed to projected attenuation matrix $\boldsymbol{P}$ by element-wise approximation (3.6). The benefit of using $\boldsymbol{P}$ is the exact linear relationship between $\boldsymbol{P}$ and $\boldsymbol{A}$, however the accuracy of $\boldsymbol{P}$ should be further considered.

For multi-energy reconstruction, data in both projection domain and image domain are organized in matrix form. This is different from single-energy models in Chapter 2, where image and sinogram are organized in vector form. Therefore, in order to formulate an objective function, two matrix operations, Kronecker product denoted by $\otimes$, and vectorization denoted by $\mathrm{vec}(\cdot)$, are used in this thesis and they are defined as below

$$\boldsymbol{C}_{mp \times nq} = \boldsymbol{A}_{m \times n} \otimes \boldsymbol{B}_{p \times q} = \begin{bmatrix} a_{11}\boldsymbol{B} & a_{12}\boldsymbol{B} & \dots & a_{1n}\boldsymbol{B} \\ a_{21}\boldsymbol{B} & a_{22}\boldsymbol{B} & \dots & a_{2n}\boldsymbol{B} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}\boldsymbol{B} & a_{m2}\boldsymbol{B} & \dots & a_{mn}\boldsymbol{B} \end{bmatrix}, \tag{3.14}$$

$$\mathrm{vec}(\boldsymbol{A}_{m \times n}) = [a_{11}, a_{21}, \dots, a_{m1}, a_{12}, a_{22}, \dots, a_{m2}, \dots, a_{1n}, a_{2n}, \dots, a_{mn}]^T. \tag{3.15}$$

Now consider formulating the simplest problem for multi-energy reconstruction when no regularization term is added and no data statistics is considered. Using $\boldsymbol{P}$ as input data and the linear relationship (3.10) and (3.12), $\boldsymbol{A}$ can be obtained by solving the following least-squares problem

$$\boldsymbol{A}^* = \arg\min_{\boldsymbol{A}} \|\mathrm{vec}(\boldsymbol{W}\boldsymbol{A}\boldsymbol{F} - \boldsymbol{P})\|_2^2 = \arg\min_{\boldsymbol{A}} \left\| \left(\boldsymbol{F}^T \otimes \boldsymbol{W}\right) \mathrm{vec}(\boldsymbol{A}) - \mathrm{vec}(\boldsymbol{P}) \right\|_2^2. \tag{3.16}$$

When $\boldsymbol{A}$ and $\boldsymbol{P}$ are vectorized, the previous problem has the same form as the single-energy least-squares problem (2.15). Consider an overdetermined system with $N_e > N_b$ and $N_p > N_i$, and exact solution for (3.16) is

$$\begin{aligned} \mathrm{vec}(\boldsymbol{A}^*) &= \left(\left(\boldsymbol{F}^T \otimes \boldsymbol{W}\right)^T \left(\boldsymbol{F}^T \otimes \boldsymbol{W}\right) \right)^{-1} \left(\boldsymbol{F}^T \otimes \boldsymbol{W}\right)^T \mathrm{vec}(\boldsymbol{P}) \\ &= \mathrm{vec}\left(\left(\boldsymbol{W}^T\boldsymbol{W}\right)^{-1} \boldsymbol{W}^T \boldsymbol{P} \boldsymbol{F}^T \left(\boldsymbol{F}\boldsymbol{F}^T\right)^{-1} \right). \end{aligned} \tag{3.17}$$

Simplification of the solution in (3.17) uses properties (510-512, 514) in [40].

3.2.3 Two-step algorithms

Problem (3.13) can be split into two steps in two different ways:

$$\begin{aligned} \boldsymbol{U}^* &= \arg\min_{\boldsymbol{U}} J_{IU}(\boldsymbol{U}; \boldsymbol{W}, \boldsymbol{I}), \\ \boldsymbol{A}^* &= \arg\min_{\boldsymbol{A}} J_{UA}\left(\boldsymbol{A}; \boldsymbol{F}, \boldsymbol{U}^*\right); \end{aligned} \tag{3.18}$$

or

$$\begin{aligned} \boldsymbol{Y}^* &= \arg\min_{\boldsymbol{Y}} J_{IY}(\boldsymbol{Y}; \boldsymbol{F}, \boldsymbol{I}), \\ \boldsymbol{A}^* &= \arg\min_{\boldsymbol{A}} J_{YA}\left(\boldsymbol{A}; \boldsymbol{W}, \boldsymbol{Y}^*\right). \end{aligned} \tag{3.19}$$

These two models summarize two different types of algorithms: (3.18) solves inverse radon transform first, and then conducts decomposition in image domain; whereas (3.19) first conducts decomposition in projection domain, and solves the inverse radon transform problem afterwards.

In general, two-step models (3.18) and (3.19) are not equivalent to the one-step model (3.13). However, formulate a two-step least-squares problem in form of (3.18) as

$$\begin{aligned} \boldsymbol{U}^* &= \arg\min_{\boldsymbol{U}} \|\mathrm{vec}(\boldsymbol{W}\boldsymbol{U} - \boldsymbol{P})\|_2^2, \\ \boldsymbol{A}^* &= \arg\min_{\boldsymbol{A}} \|\mathrm{vec}\left(\boldsymbol{A}\boldsymbol{F} - \boldsymbol{U}^*\right)\|_2^2, \end{aligned} \tag{3.20}$$

and the solution for overdetermined system with $N_e \geq N_b$ and $N_p \geq N_i$ is given by

$$\begin{aligned} \boldsymbol{U}^* &= \left(\boldsymbol{W}^T\boldsymbol{W}\right)^{-1}\boldsymbol{W}^T\boldsymbol{P}, \\ \boldsymbol{A}^* &= \boldsymbol{U}^*\boldsymbol{F}^T\left(\boldsymbol{F}\boldsymbol{F}^T\right)^{-1} = \left(\boldsymbol{W}^T\boldsymbol{W}\right)^{-1}\boldsymbol{W}^T\boldsymbol{P}\boldsymbol{F}^T\left(\boldsymbol{F}\boldsymbol{F}^T\right)^{-1}. \end{aligned} \tag{3.21}$$

It can be seen that (3.16) and (3.20) have the same solution. Similarly, a least-squares problem formulated in form of (3.19) should also have the same solution.

3.3 Formulation of a two-step algorithm

In this section, a two-step algorithm in form of (3.19) is formulated. The first step of the algorithm constructs and solves a maximum-likelihood model by considering photon statistics, and conducts attenuation decomposition in projection domain. The second step formulates and solves a penalized weighted least-squares (PWLS) problem to obtain images of coefficients. The first step of the algorithm was originally reported in [41], and the second step in [42, 43].

3.3.1 Maximum-likelihood algorithm for projection data decomposition

Let $\boldsymbol{I}_k^{row} = [I_{k,1}, I_{k,2}, ..., I_{k,N_e}]$ be the kth row vector of matrix $\boldsymbol{I}$. Considering Poisson photon statistics, the conditional probability of obtaining measurement results $\boldsymbol{I}_k^{row}$ given the expected photon number $\bar{\boldsymbol{I}}_k^{row}$ can be expressed as

$$P\left(\boldsymbol{I}_k^{row} | \bar{\boldsymbol{I}}_k^{row}\right)\right) = \prod_{i=1}^{N_e} \frac{\bar{I}_{k,i}^{I_{k,i}} \cdot e^{-\bar{I}_{k,i}}}{I_{k,i}!}. \tag{3.22}$$

Substituting $p(E, \theta, \xi)$ in (3.5) by (3.4) gives $\bar{I}_{k,i}$ as a function of $\boldsymbol{Y}_{\boldsymbol{k}}^{\boldsymbol{row}}$

$$\bar{I}_{k,i}(\boldsymbol{Y}_k^{row}) = \int_{E \in B_i} I_{in}(E) \exp\left(-\sum_{j=1}^{N_b} f_j(E) \cdot Y_{k,j}\right) D(E) dE, \tag{3.23}$$

where $\boldsymbol{Y}_{\boldsymbol{k}}^{\boldsymbol{row}}$ is the row vector of matrix $\boldsymbol{Y}$.

Taking logarithm of both sides in (3.22) and substituting $\bar{I}_{k,i}$ by (3.23) gives a log-likelihood function for $\boldsymbol{Y}_k^{row}$

$$L\left(\boldsymbol{I}_k|\boldsymbol{Y}_k^{row}\right) = \sum_{i=1}^{N_e}\left(I_{k,i}\ln(\bar{I}_{k,i}(\boldsymbol{Y}_k^{row})) - \bar{I}_{k,i}(\boldsymbol{Y}_k^{row})\right). \tag{3.24}$$

Maximizing (3.24) gives the estimated $(\boldsymbol{Y}_k^{row})^*$ with maximal probability. Optimal matrix $\boldsymbol{Y}^*$ can be obtained by solving N_p parallelizable optimization problems by using the downhill simplex method [44]. In general $L\left(\boldsymbol{I}_k|\boldsymbol{Y}_k^{row}\right)$ in (3.19) is not a concave function.

Appoximately pulling exponential parts out of the integral in equation (3.23) gives

$$\bar{I}_{k,i}(\boldsymbol{Y}_k^{row}) \approx \exp\left(-\langle\boldsymbol{Y}_k^{row}, \boldsymbol{f}_i\rangle\right)\int_{E\in B_i} I_{in}(E)D(E)dE, \tag{3.25}$$

where $\boldsymbol{f}_i$ is the ith column vector in matrix $\boldsymbol{F}$. Substituting $\bar{I}_{k,i}$ now by (3.25) gives an approximated log-likelihood function

$$\widetilde{L}\left(\boldsymbol{I}_k|\boldsymbol{Y}_k^{row}\right) = \sum_{i=1}^{N_e}\left(-I_{k,i}\langle\boldsymbol{Y}_k^{row}, \boldsymbol{f}_i\rangle - \exp\left(-\langle\boldsymbol{Y}_k^{row}, \boldsymbol{f}_i\rangle\right)\int_{E\in B_i} I_{in}(E)D(E)dE\right). \tag{3.26}$$

Note that after the approximation, (3.26) is similar to likelihood function (2.18) in Section 2.3, and it is a concave function of $\boldsymbol{Y}_k^{row}$. Therefore, a similar weighted least-squares problem can be formulated following the same routine in Section 2.3. Therefore , a weighted least-squares problem is formulated as

$$\boldsymbol{Y}_k^{row,*} = \arg\min_{\boldsymbol{Y}_k^{row}}\left\{\|\boldsymbol{Y}_k^{row}\boldsymbol{F} - \boldsymbol{P}_k^{row}\|_{C_k}^2\right\}, \tag{3.27}$$

where C_k is a diagonal matrix with entries $I_{k,1}, I_{k,2}, ..., I_{k,N_e}$, and it is the inverse of covariance matrix for projection vector $\boldsymbol{P}_k^{row}$.

3.3.2 Reconstructing material decomposed sinogram

As the second step of the algorithm, a regularized weighted least-squares problem is formulated as below

$$\boldsymbol{A}^* = \arg\min_{\boldsymbol{A}}\left(\|\mathrm{vec}\left(\boldsymbol{W}\boldsymbol{A} - \boldsymbol{Y}^*\right)\|_\Sigma^2 + \sum_{i=1}^{N_b}\alpha_i J_{Regu}(\boldsymbol{A_i})\right), \tag{3.28}$$

where $\boldsymbol{A_i}$ is the ith column vector of matrix $\boldsymbol{A}$ and it is the decomposition coefficient image for the ith basis. $J_{Regu}(\boldsymbol{A_i})$ is a regularization term for image vector $\boldsymbol{A_i}$, and it can use the same definition as equation (2.23) and (2.31) in Chapter 2, and α_i is the corresponding regularization parameter. Weight matrix for the data fidelity term Σ can be defined as the inverse of the covariance matrix of vec($\boldsymbol{Y}^*$). Note that matrix $\boldsymbol{Y}^*$ is solved row by row from previous step, and covariance for vec($\boldsymbol{I}$) as the input for step-one problem is a diagonal matrix. Therefore, there will be only column correlation but no row correlation in matrix $\boldsymbol{Y}^*$, which means for the ith column vector of matrix $\boldsymbol{Y}^*$, $cov(\boldsymbol{Y}_i^*)$ is a diagonal matrix but for the jth row vector $\boldsymbol{Y}_j^{row,*}$, $cov(\boldsymbol{Y}_j^{row,*})$ is a full matrix of size $N_b \times N_b$.

In order to formulate a problem with weight matrix of simpler form, the data fidelity term is transformed to obtain the following problem

$$\boldsymbol{A}^* = \arg\min_{\boldsymbol{A}}\left(\left\|\mathrm{vec}\left((\boldsymbol{W}\boldsymbol{A})^T\right) - \mathrm{vec}\left((\boldsymbol{Y}^*)^T\right)\right\|_{\Sigma_T}^2 + \sum_{i=1}^{N_b}\alpha_i J_{Regu}(\boldsymbol{A_i})\right), \tag{3.29}$$

where inverse of Σ_T is given by

$$\Sigma_T^{-1} = cov(\text{vec}\left((\boldsymbol{Y}^*)^T\right)) = \text{diag}\left\{cov(\boldsymbol{Y}_1^{row,*}), cov(\boldsymbol{Y}_2^{row,*}), ..., cov(\boldsymbol{Y}_{N_p}^{row,*})\right\}. \tag{3.30}$$

Σ_T^{-1} is a block diagonal matrix, and thus Σ_T is also a block diagonal matrix given by

$$\Sigma_T = \text{diag}\left\{\left[cov(\boldsymbol{Y}_1^{row,*})\right]^{-1}, \left[cov(\boldsymbol{Y}_2^{row,*})\right]^{-1}, ..., \left[cov(\boldsymbol{Y}_{N_p}^{row,*})\right]^{-1}\right\}. \tag{3.31}$$

Numerical methods discussed in previous Chapter 2, such us quasi-Newton and ADMM, now can be used to solve optimization problem (3.29).

Derivation of $cov(\boldsymbol{Y}_k^{row,*})$

Now we need to compute $cov(\boldsymbol{Y}_k^{row,*})$ as blocks in Σ_T^{-1}. The following theorem explains how to compute covariance of maximum-likelihood estimators.

Theorem 3 *(Frechet–Darmois–Cramér–Rao inequality, Section 6.6 in [45])*
For multivariate maximum-likelihood estimator $\hat{\boldsymbol{\theta}} = L(\boldsymbol{\theta})$, the inverse of the covariance matrix of their estimators $V_{i,j} = cov\left(\widehat{\theta_i}, \widehat{\theta_j}\right)$ is given by (assuming efficiency and zero bias)

$$V_{i,j}^{-1} = E\left[-\frac{\partial^2 \log L}{\partial\theta_i \partial\theta_j}\right], \tag{3.32}$$

where $E[\cdot]$ computes the expectation.

To simplify computation, the approximated expression for $\bar{I}_{k,i}(\boldsymbol{Y}_k^{row})$ (3.25) and the corresponding approximated maximum likelihood function (3.24) is used for second order derivative computation,

$$\begin{aligned}
\frac{\partial L\left(\boldsymbol{I}_k|\boldsymbol{Y}_k^{row}\right)}{\partial Y_{k,i}} &= \sum_{m=1}^{N_e} \frac{\partial L\left(\boldsymbol{I}_k|\boldsymbol{Y}_k^{row}\right)}{\partial \bar{I}_{k,m}\left(\boldsymbol{Y}_k^{row}\right)} \cdot \frac{\partial \bar{I}_{k,m}\left(\boldsymbol{Y}_k^{row}\right)}{\partial Y_{k,i}} \\
&= \sum_{m=1}^{N_e} \left(\frac{I_{k,m}}{\bar{I}_{k,m}\left(\boldsymbol{Y}_k^{row}\right)} - 1\right) \cdot \left(-f_{m,i} \cdot \bar{I}_{k,m}\left(\boldsymbol{Y}_k^{row}\right)\right) \\
&= \sum_{m=1}^{N_e} \left(-I_{k,m} \cdot f_{m,i} + f_{m,i} \cdot \bar{I}_{k,m}\left(\boldsymbol{Y}_k^{row}\right)\right),
\end{aligned} \tag{3.33}$$

$$\frac{\partial^2 L\left(\boldsymbol{I}_k|\boldsymbol{Y}_k^{row}\right)}{\partial Y_{k,j}\partial Y_{k,i}} = \sum_{m=1}^{N_e} \left(f_{m,i} \cdot \left(-\sum_{n=1}^{N_e} f_{n,i} \cdot \bar{I}_{k,n}\left(\boldsymbol{Y}_k^{row}\right)\right)\right). \tag{3.34}$$

Summarizing the above result (3.34) in a matrix form gives the covariance estimation

$$\begin{aligned}
\left[cov(\boldsymbol{Y}_k^{row,*})\right]^{-1} &= \left[\frac{\partial^2 L\left(\boldsymbol{I}_k|\boldsymbol{Y}_k^{row}\right)}{\partial Y_{k,j}\partial Y_{k,i}}\right]_{i,j=1,...,N_b} = \boldsymbol{F}^T \cdot \text{diag}\left\{\bar{I}_{k,1}, ..., \bar{I}_{k,N_e}\right\} \cdot \boldsymbol{F} \\
&\approx \boldsymbol{F}^T \cdot \text{diag}\left\{I_{k,1}, ..., I_{k,N_e}\right\} \cdot \boldsymbol{F}.
\end{aligned} \tag{3.35}$$

In the above equation (3.35), measurements $I_{k,1}, ..., I_{k,N_e}$ are used for estimating expectation $\bar{I}_{k,1}, ..., \bar{I}_{k,N_e}$.

Label	Material	Density (kg m^{-3}) in 20 °C
1	soft tissue	1020
2	heart (blood filled)	1060
3	yellow marrow	980
4	cortical bone	1920
5	contrast agent + water ($m^{\%}_{Gd} = 3\%$)	1061
6, 7	contrast agent + water ($m^{\%}_{Gd} = 1\%$)	1019

Table 3.1: Material types used in phantom simulations [25, 37].

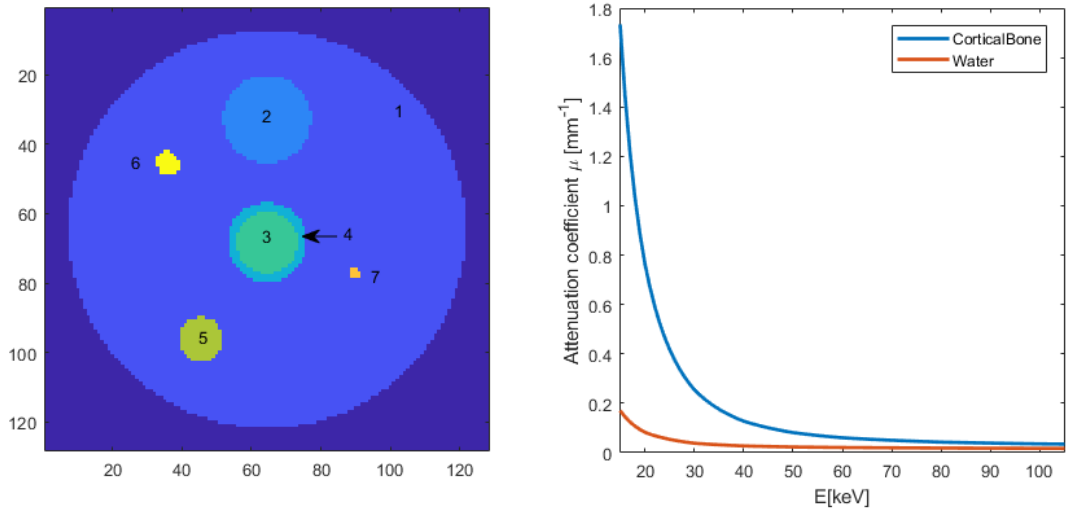
3.4 Numerical experiments

3.4.1 Experimental setups

In general, experimental setups are similar to that in the single-energy experiment in Section 2.5.1. The uniqueness is how to setup a phantom including K-edge contrast agent, and how to generate spectral noisy sinograms.

Phantom description

A simulated phantom including 7 different regions is shown in Figure 3.3a, and material type description is summarized in Table 3.1. Region 1-4 are body tissues, and Figure 3.3b shows that attenuation coefficient curves of cortical bone and water versus energy differ greatly. Region 5-7 are a mixture of water and a gadolinium (Gd) -based contrast agent. In this thesis, a contrast agent with chemical formula $C_{16}H_{31}GdN_5O_8$ is simulated in different concentrations in water. For high concentration region 5, the simulated gadolinium concentration in mass percentage, $m^{Gd}_{\%}$, is 3%; for low concentration regions 6 and 7, $m^{Gd}_{\%}$ is 1%. Local density of contrast region 5-7 can be computed following the method given in [41, Appendix equation A.1 and A. 4]. The dimensions of the phantom are 300×300 mm^2, consisting of 256×256 pixels.



(a) True phantom labeling 7 different regions. (b) MAC curves for two different body tissues.

Figure 3.3: Phantom description for multi-energy experiments.

CT geometry and system matrix $\boldsymbol{W}$

CT geometry for multi-energy experiments is similar as in Chapter 2, and the same method introduced is used to generate system matrix $\boldsymbol{W}$. For experiments in this chapter, 336 equally spacing channels are measured on a 300mm-long detector, and 180 projections are measured in the 2π rotating range.

Generation of multi-energy sinogram

For multi-energy experiments, nine energy bins are selected, $B_1 = [15, 25]$ keV, $B_2 = [25, 35]$ keV, ..., $B_9 = [95, 105]$ keV. Algorithm 6 used for generation of noisy spectral data is similar to Algorithm 5 for single-energy cases, in which noiseless projection data is converted to photon numbers in order to apply the Poisson noise model. Note that input of Algorithm 6, x-ray tube spectra $I_{in}(E)$ and clean projection data $\boldsymbol{p}_{clean}(E)$ are both functions of energy E. Input x-ray spectra $I_{in}(E)$ is plotted in Figure 3.4, and the total number of input photons within energy range [15, 105] keV for each channel is $1.03 * 10^6$. Two attenuation images at different keVs and the corresponding clean projection data is shown in Figure 3.5. It is shown that attenuation at 47keV and 53keV differs greatly in contrast regions.

For multi-energy simulation, the attenuation coefficients are much larger in the lower energy range, therefore the photon number simulated after step 3 can be 0 for many channels. When the photon number $I_{k,j}$ is 0, it is impossible to compute sinogram $p_{k,j}$ in step 4, and such phenomenon is called photon starvation. If sinogram data is still needed, a minimum threshold for photon number should be set to ensure $I_{k,j} > 0$ for all k and j.

Algorithm 6 Generate noisy data $\boldsymbol{p_j}$ and $\boldsymbol{I}_j$ for energy bin $B_j = [E_j^{low}, E_j^{high}]$.

1: Input: $I_{in}(E)$, x-ray source photon number; $\boldsymbol{p}_{clean}(E)$, simulated clean sinogram;
2: Compute expected photon number $\bar{I}_{k,j}$ by equation (3.5) for $k = 1, 2, ..., N_p$;
3: Compute measured photon number with Poisson noise:

$$I_{k,j} = \text{Poisson}(\bar{I}_{k,j}), \text{ for } k = 1, 2, ..., N_p;$$

4: Compute noisy sinogram $p_{k,j}$ by equation (3.6) for $k = 1, 2, ..., N_p$;

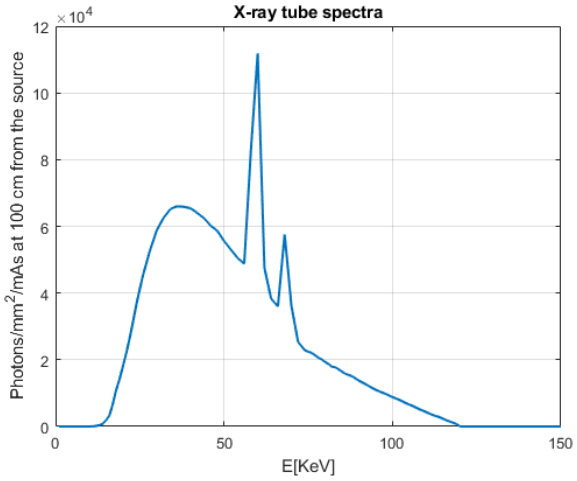


Figure 3.4: X-ray tube spectra calculated by SPEKTR 3.0 toolbox [46]. TASMIP spectrum with 1.6 mm Al inherent filtration and 0% kV ripple.

(a) True attenuation Image at 47keV.

(b) True attenuation Image at 53keV.

(c) Attenuation at projection location 1.

Figure 3.5: Clean attenuation images and projections at different keVs.

Tested models and algorithms

The two-step algorithm developed in previous Section 3.3 is experimented here. For the first step, three basis functions are selected for sinogram decomposition: f_1 defined by (3.2) models the photoelectric absorption effect, f_2 defined by (3.3) models the Compton scattering effect, and f_3 is the MAC for Gadolinium element. Therefore, image reconstructed from decomposition coefficients for the third basis is the local density (unit in g/cm^3) of the K-edge element.

For the second step, the weighted least-squares problem (3.29) formulated in Section 3.3.2 is solved by quasi-Newton algorithm discussed in Chapter 2. Specifically, the preconditioning technique is applied to accelerate the convergence, and the problem is solved without or with different regularization terms.

3.4.2 Results

Sinogram decomposition

Results of step one algorithm (sinogram decomposition) are summarized in Figure 3.6. Two different models, the approximated least-squares model (3.27) and the exact maximum-likelihood model (3.24), are experimented. For (3.27) using all spectral data, decomposed sinogram for the 3rd (contrast agent) basis is shown in Figure 3.6a, and the absolute error comparing to true sinogram is shown in Figure 3.6b. In the true sinogram for contrast agent basis, there should be three sine-like curves because of the three contrast agent regions in the image. It can be observed that decomposition errors are large at many center channels. This is caused by photon starvation of low energy sinograms.

Instead of using all sinograms measured by 9 energy bins, model (3.27) is solved again using sinograms measured by 6 energy bins, [45,55], [55,65], ..., [95,105]keV. Because sinograms at energy bins [15,25], [25,35], [35,45] keV suffer from photon starvation and are thrown away now, decomposition results are much more accurate comparing to using all spectral sinograms, as shown in Figure 3.6c and 3.6d. Mean absolute error for the 3rd decomposed sinogram is $3.33 * 10^{-3}$.

Different from (3.27), model (3.24) is an exact maximum-likelihood model and uses photon measurements $\boldsymbol{I}$ as input instead of using sinograms $\boldsymbol{P}$. Since (3.24) is a non-concave function, proper initial points should be chosen in order to find the global maximums. (3.24) is firstly solved by using all 0 initial points, and results are shown in Figure 3.6e and 3.6f. It can be observed that some pixels in the sinogram are not corrected decomposed.

Since model (3.27) using partial spectral data gives results with no visible errors, these results will now be used as initial points for solving the exact model (3.24). Results using

proper initial points are shown in Figure 3.6g and 3.6h. Mean absolute error now is $2.50 * 10^{-3}$. Figure 3.6 and the computed absolute errors demonstrate that solving the exact model (3.24) by using approximated model (3.27) results as initial points is the best tested method for sinogram decomposition, and results by this method will be used for the step two experiments.

Reconstructing decomposed sinogram

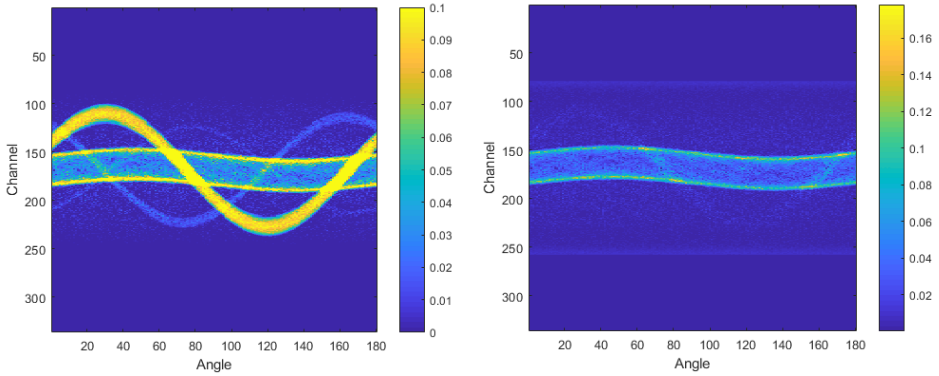
Results with no regularization To test the step two method in Section 3.3.2 and especially the estimated inverse covariance matrix Σ_T, an iterative reconstruction is conducted using fully estimated inverse covariance and no regularization term, and the results are presented in Figure 3.7a-3.7c. The three reconstructed images are coefficients for absorption, scattering and contrast element (Gd) concentration respectively. FBP results are presented in Figure 3.7d-3.7f, and it is observed that iterative reconstruction (IR) results are similar but noisier comparing to FBP results.

For material decomposed images, only the Gd concentration image can be compared to true image, becuase 'true' absorption or 'true' scattering image can not be obtained. After reconstructing the material decomposed images, attenuation images at different energies can also be synthesized by linear relationship (3.1). Attenuation images at 60keV computed from IR results are FBP results are presented in Figure 3.8, and a true image is also given for comparison. It is shown that IR attenuation image 3.8a is more accurate than the FBP attenuation image 3.8b, especially in the bone region.

One noticeable observation during the step two experiment is the extremely slow convergence after incorporating estimated inverse covariance. The preconditioning technique developed in Chapter 2 (see Page 10) should also be used here. When no regularization is used, 200 iterations of BFGS quasi-Newton with steepest descent step size are run, and reconstruction results with or without preconditioning are shown in Figure 3.9. It shows that both objective function and error norm converge much faster when applying preconditioning. After 200 iterations, no-precondition Gd concentration image 3.9c shows severe artifacts near bone regions and outer boundaries, however only the three contrast regions and random noises show up in the precondition image 3.9d. The inter-sinogram correlations, which are the off-diagonal entries of the inverse covariance matrix Σ_T, make the model even more ill-conditioned. For the current experiment, the condition number of Σ_T is $3.780 * 10^5$. If all inter-sinogram correlations are removed, the original problem (3.28) can be split into three sub-problems for reconstruction $\boldsymbol{A}_1$,$\boldsymbol{A}_2$ and $\boldsymbol{A}_3$ respectively, and condition number for the splitted inverse covariance matrices are 5266.1, 20.80 and 143.75 , all of which much smaller than the original one. The numerical experiments demonstrate the necessity of applying preconditioning technique for reconstructing material-decomposed sinograms with full inverse covariance estimation.

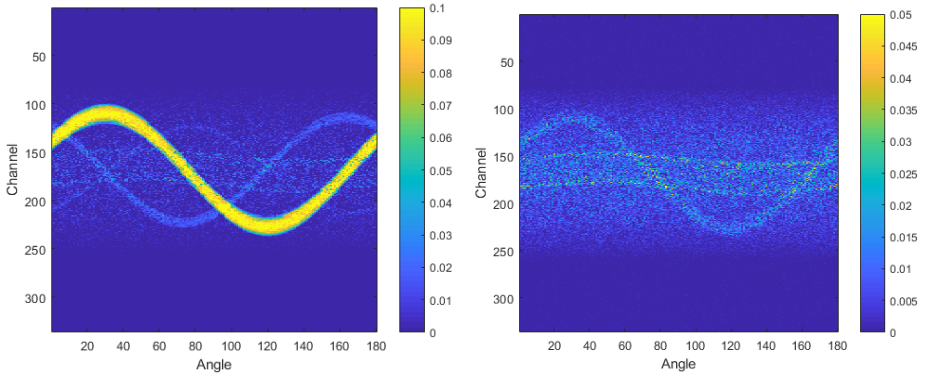
Removing inter-sinogram correlations Since reconstructing the material basis images with full inverse correlation matrix yields very slow convergence, a diagonal inverse covariance matrix diag$\{\Sigma_T\}$ is used now and the basis images can be reconstructed separately from each decomposed sinograms. Experiments are conducted to compare reconstructions using only diagonal inverse covariance matrix as weight matrix with using no weight matrix. Since the step two problem can be split into three independent problems after removing inter-sinogram correlations, only the contrast agent basis image is reconstructed, and the results are summarized in Figure 3.10.

Figure 3.10a shows the resolution-noise tradeoff curves for reconstructions using TV regularization, and it is seen that TV-regularized reconstruction has lower noise level as well as higher resolution when using diagonal inverse covariance matrix. Figure 3.10b and 3.10c are reconstructed images using Tikhonov regularization. Figure 3.10b uses diagonal inverse covariance matrix as weight matrix and regularization parameter α_3 is 3. Figure 3.10c uses no weight



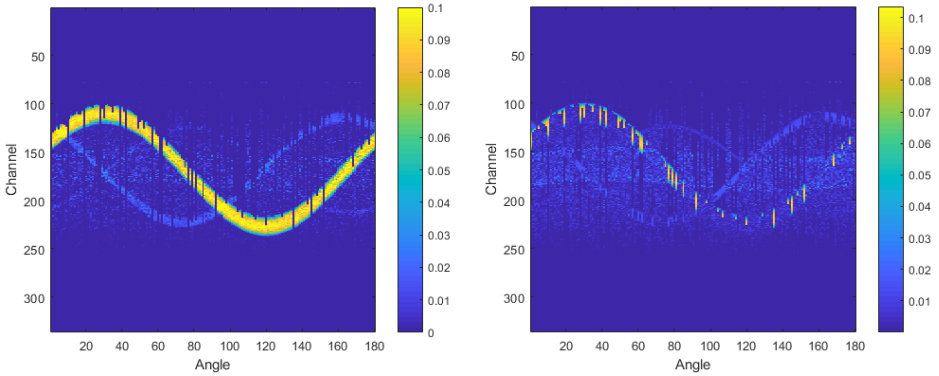
(a) Model (3.27), data in [15,105] keV

(b) |(a) - True|



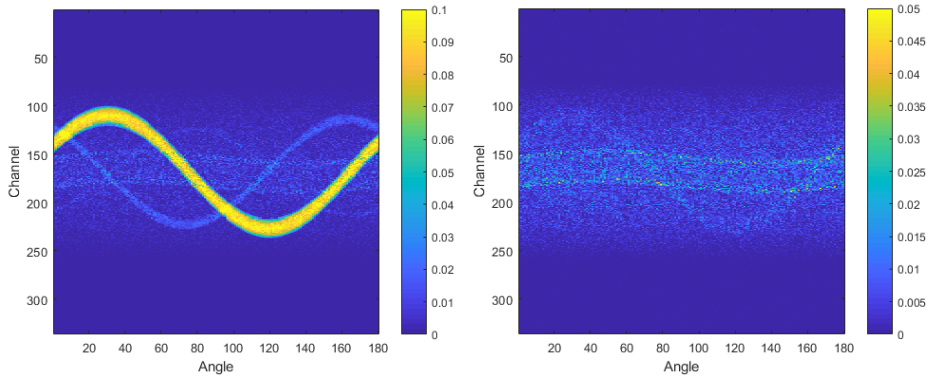
(c) Model (3.27), data in [45,105] keV

(d) |(c) - True|



(e) Model (3.24), use all 0 initial points

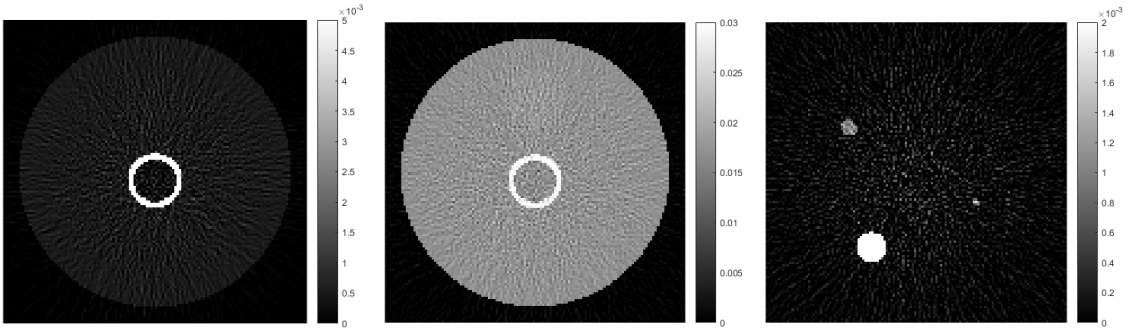
(f) |(e) - True|



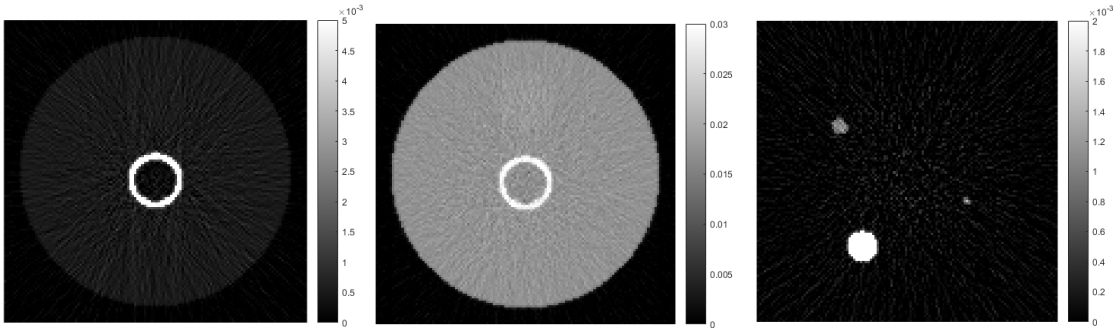
(g) Model(3.24), use (c) as initial points

(h) |(g) - True|

Figure 3.6: Decomposed sinogram for the contrast agent basis. (a),(c),(e) and (g) are results of four different experiments; (b),(d),(f) and (h) are the corresponding absolute error figures.

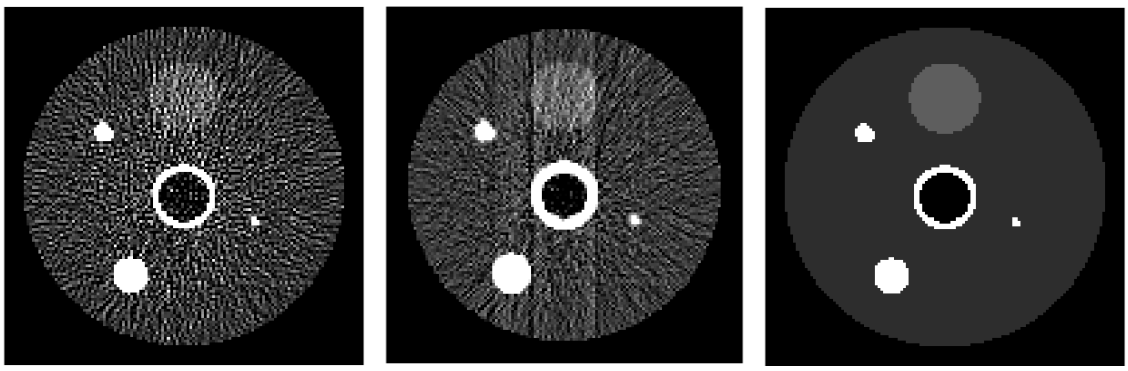


(a) Iterative, absorption (b) Iterative, scattering (c) Iterative, contrast



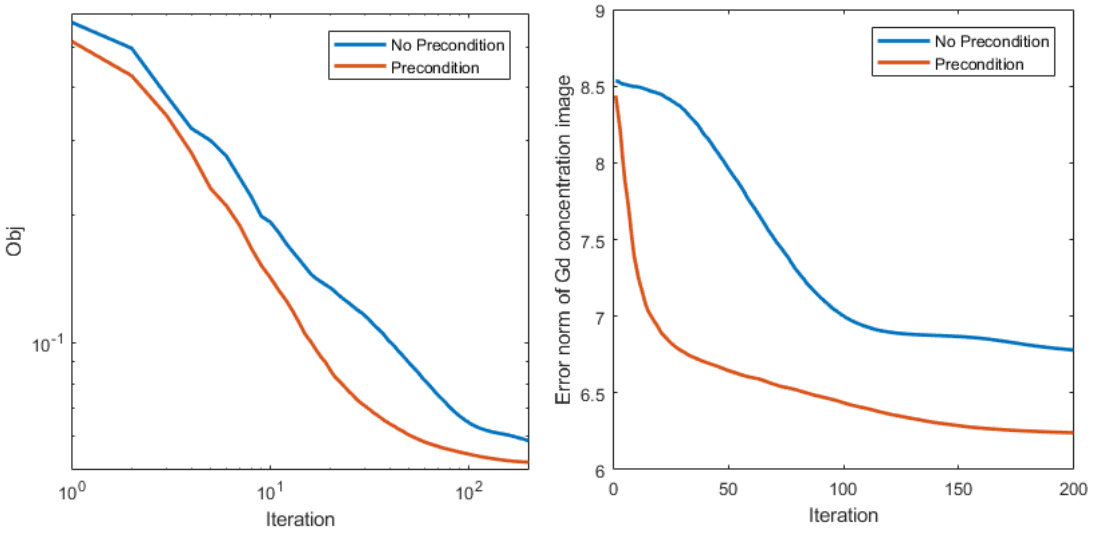
(d) FBP, absorption (e) FBP, scattering (f) FBP, contrast

Figure 3.7: Decomposed sinogram reconstructions. Iterative reconstruction (up) with no regularization and fully estimated covariance. FBP (down) reconstruction for comparison.



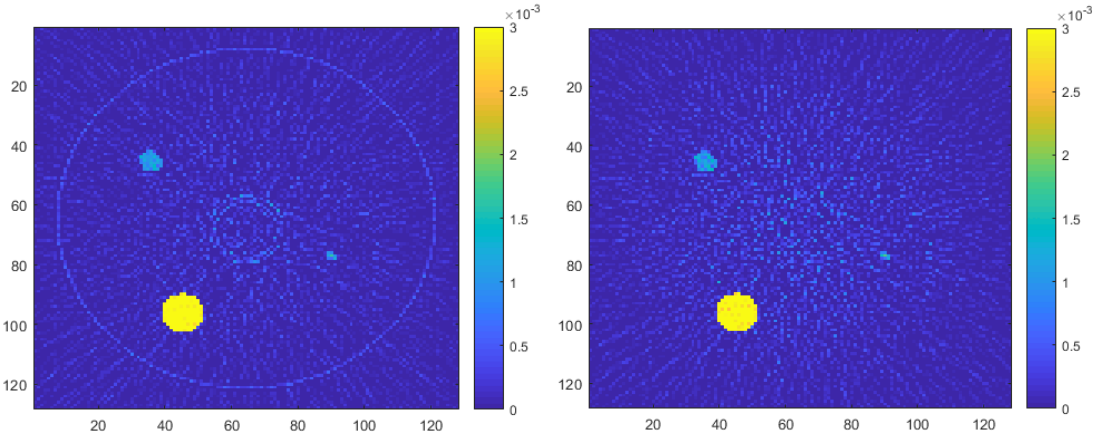
(a) Iterative (b) FBP (c) True

Figure 3.8: Attenuation coefficient images at 60keV. (a) is computed from Figure 3.7a-3.7c, and (b) is computed from Figure 3.7d-3.7f. Display window $[0.02\ 0.025]\text{mm}^{-1}$.



(a) Obj vs Iteration (log-log)

(b) $\|\boldsymbol{A}_3^* - \boldsymbol{A}_{True}\|_2$ vs Iteration



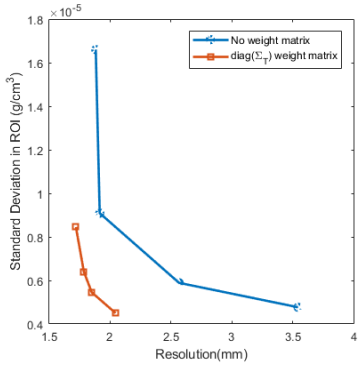
(c) 200 iterations, no precondition.

(d) 200 iterations, precondition.

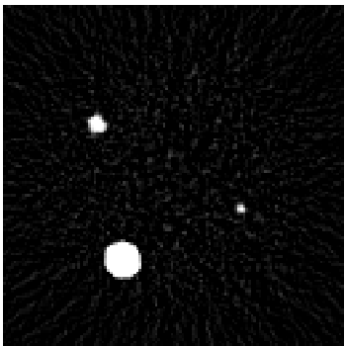
Figure 3.9: Convergence analysis for no regularization experiment. Comparison between precondition and no-precondition .

matrix and regularization parameter α_3 is 20. The selection of different α_3 values is to ensure the overall error and the resolution of the two images to be similar. It can be observed that the noise level becomes lower especially in the central region of the image and the streak artifacts are suppressed after applying diagonal inverse covariance matrix.

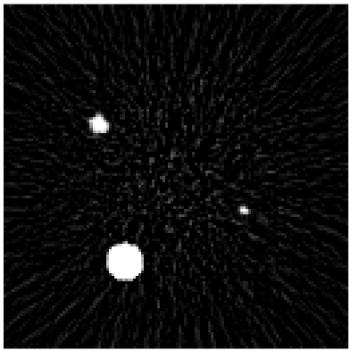
In general, it can be concluded that material-specific reconstruction yields better results when using diagonal inverse covariance matrix $\text{diag}\{\Sigma_T\}$ comparing to using no weight matrix for data fidelity term. However, how the inter-sinogram correlations will effect the reconstruction results has not been fully studied yet.



(a) Resolution-noise tradeoff curves for TV regularization.



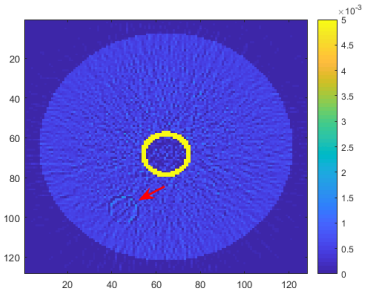
(b) Tikhonov regularization, $\alpha = 3$, Σ_T weight matrix.



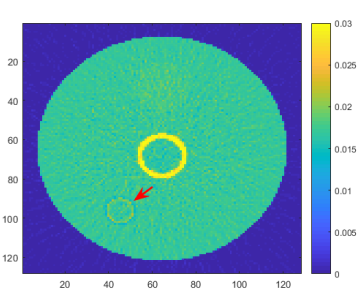
(c) Tikhonov regularization, $\alpha = 20$, no weight matrix.

Figure 3.10: Comparison between using $\text{diag}\{\Sigma_T\}$ and using identity matrix as weighted matrix.

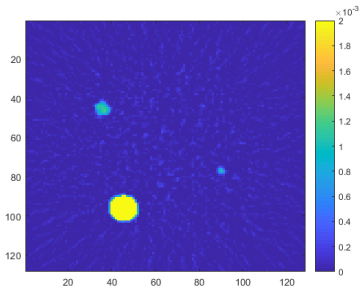
Choosing regularization parameters when using Σ_T When reconstructing material decomposed sinograms simultaneously using model (3.29), there will be N_b (in the experiment $N_b = 3$) regularization parameters $\alpha_1, ..., \alpha_{N_b}$ used, each one for regularizing one material basis image, and no inter-material correlations will be modeled in the regularization terms. When reconstructing one sinogram as the single-energy case in Chapter 2, L-curve may serves as a simple way to choose the proper regularization parameter. However when reconstructing multiple sinograms in one model and the inter-sinogram correlations are modeled by weight matrix Σ_T, how to properly choose regularization parameters for every reconstructed image remains a problem. When regularization term for the ith image is over-weighted, the difference between the over-regularized image and the true image will cause artifacts in other images, as shown in Figure 3.11. It can be observed that over-regularized contrast agent image will cause artifacts in the high contrast region of the other two images, however when no regularization is used for all the three images, such artifacts can not be observed, as shown in Figure 3.9a-3.9c.



(a) Absorption image.



(b) Scattering image.



(c) Contrast agent image.

Figure 3.11: Reconstructing material decompsed sinograms with regularization parameter $\boldsymbol{\alpha} = [0, 0, 0.01]^T$. TV regularization used for model (3.29).

3.4.3 Discussion and conclusion

In Section 3.4, a two-step algorithm developed in Section 3.3 for spectral CT reconstruction is tested using simulated noisy spectral data. For the sinogram decomposition step, two models, the exact maximum-likelihood ML) model (3.24) and the approximate weighted least-squares (LS) model (3.27), are both experimented. It use shown that the exact ML model yields more accurate decomposition result, however it requires proper initial point to achieve the global maximum. In the original reference [41] there is no detailed discussion about how to properly solve the non-concave ML problem. In the experiments in this thesis, it is shown that the result of LS model can be used as a proper initial point for solving the exact ML model by the downhill simplex method. How to ensure the global optimal solution is found requires further theoretical analysis of the ML function (3.24).

For the second step of the two-step algorithm, it is shown that the regularized weighted least-squares model (3.29) can be used to reconstruct the material basis images. Experiments show that numerical algorithms developed in Chapter 2 can also be used for solving the reconstruction problem for multi-sinogram cases. Experimental results also show that modeling the variance of the decomposed sinogram in the weight matrix yields better reconstruction results comparing to using no weight matrix. However a fully estimated inverse covariance matrix used as the weight matrix will make the problem even more ill-conditioned and more difficult to solve, and the effect of modeling inter-sinogram correlation in the weight matrix needs further investigation as well.

Chapter 4

Discussion, conclusions and future work

4.1 Discussion and conclusion

In this thesis, reconstruction algorithms for spectral computed tomography are studied and tested. A reconstruction algorithm for inverse problem usually consists of: (1) formulating an optimization problem that includes a data fidelity term and a regularization term; (2) applying a proper numerical method to solve the optimization problem.

As a starting point of this thesis, CT reconstruction algorithms for single-energy cases are firstly studied in Chapter 2. It is proved that the weighted least-squares problem can be approximated from a maximum-likelihood problem modeling photon statistics. It is also shown by experiments that such a weighted least-squares problem with TV or Tikhonov regularization can be solved by quasi-Newton method together with proper acceleration techniques. Such a numerical algorithm can also be applied to multi-energy cases.

When turning to multi-energy reconstruction, not only the attenuation images but also the material decomposed images can be reconstructed when specifying the material basis. In Chapter 3, the spectral reconstruction problem is formulated by modeling the CT geometry as well as the material decomposition process. In order to simplify the computation, two-step algorithms are also formulated which solve the inverse radon transform and the material decomposition separately. A two-step algorithm which solves the material decomposition in the projection domain and reconstructs the decomposed sinograms afterwards is developed and tested in Chapter 3. Simulated numerical experiments show that the material decomposition step can be conducted by solving N_p non-concave optimization problems after applying proper intial points, where N_p denotes the number of pixels in the projection domain for one energy bin. Experiments also show that material-decomposed sinograms can be reconstructed by applying the reconstruction algorithms developed in Chapter 2, and that reconstruction results have better image quality after modeling the estimated sinogram variance. All Matlab code for numerical experiments are uploaded on Github: `https://github.com/HannaLiu18/SpectralRecon`.

4.2 Possible future works

Due to time limitation of the project, several interesting problems related to CT reconstruction and especially spectral CT reconstruction which are not mentioned in previous chapters are briefly discussed here. These might be possible research topics following after this thesis project.

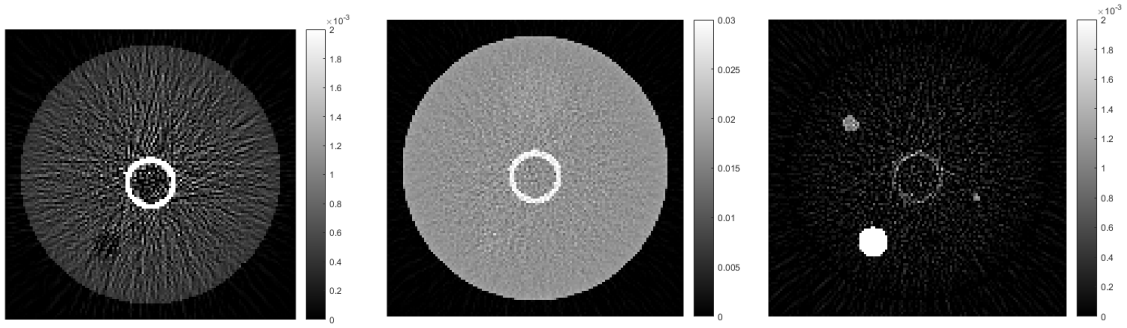
One-step algorithm for spectral reconstruction In Chapter 3, the simplest one-step model for spectral reconstruction is formulated in (3.16). When considering photon statistics

and regularization terms, the one-step model using projected attenuation $\boldsymbol{P}$ can be given as

$$\boldsymbol{A}^* = \arg\min_{\boldsymbol{A}} \left\| \left(\boldsymbol{F}^T \otimes \boldsymbol{W}\right) \mathrm{vec}(\boldsymbol{A}) - \mathrm{vec}(\boldsymbol{P}) \right\|_{\Sigma_P}^2 + \sum_{i=1}^{N_b} \alpha_i J_{Regu}(\boldsymbol{A_i}), \tag{4.1}$$

where weight matrix Σ_P denotes the inverse of the covariance matrix of $\mathrm{vec}(\boldsymbol{P})$, and all other variables use same definition as in Chapter 3. Preliminary one-step reconstruction result with no regularization term is shown in Figure 4.1, and artifacts in the bone region can be observed in the contrast agent image. Such artifacts don't exist in the two-step reconstruction results. One possible reason for the artifacts might be the insufficient convergence even after 600 iterations. Similar as single-energy case, weight matrix Σ_P is a diagonal matrix with entries of detector photon counts for all projection pixels, and photon counts will be within a even larger range for multi-energy cases. Therefore the problem (4.1) will be more ill-conditioned comparing to single-energy problem, and efficient numerical algorithms are required to solve the one-step reconstruction algorithm.

Instead of using the projected attenuation $\boldsymbol{P}$, measured photon counts $\boldsymbol{I}$ can be directly used for modeling a one-step problem. However, such a one-step problem will have a nonconvex data fidelity term, and requires further numerical algorithm development[47, 48].



(a) Absorption image. (b) Scattering image. (c) Contrast agent image.

Figure 4.1: One-step reconstruction result with no regularization. Run 600 iterations of quasi-Newton.

Designing better regularization term In this thesis, two simple regularization terms for reconstruction problem, a Tikhonov regularization and an anisotropic TV regularization, are introduced in Section 2.4 and applied in later numerical experiments for both single-energy and multi-energy cases. These two regularization terms have their own limitations. The simple Tikhonov regularization is incapable of preserving strong edges, whereas the TV regularization usually gives over-smoothed images with staircase artifacts. Many new regularization techniques have been developed in recent years, such as total generalized variance (TGV) [49] , non-local means (NLM) [50], dictionary learning (DL) [51] and etc. These techniques may be further applied to spectral CT reconstructions in order to provide images in high quality.

Modeling non-ideal detector responses Throughout this thesis, Poisson statistics model is applied for the photon-counting detectors for both single-energy and multi-energy cases. However, Poisson statistics model may not be adequate in real experiments especially when the photon counts are ultra-low [52]. Better statistical model can be obtained by estimating the noise distribution of photon counts from experimental data and then applied to reconstruction problems.

In this thesis, the realistic spectral responses of the photon-counting detectors are also not modeled. In practice, non-ideal effects include charge sharing, K-escape, and pile-up effects, and they should also be included in the reconstruction model.

Bibliography

[1] Bruno De Man and Samit Basu. "Distance-driven projection and backprojection in three dimensions". In: *Physics in Medicine & Biology* 49.11 (2004), p. 2463.

[2] Richard Gordon, Robert Bender, and Gabor T Herman. "Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and X-ray photography". In: *Journal of theoretical Biology* 29.3 (1970), pp. 471–481.

[3] Anders H Andersen and Avinash C Kak. "Simultaneous algebraic reconstruction technique (SART): a superior implementation of the ART algorithm". In: *Ultrasonic imaging* 6.1 (1984), pp. 81–94.

[4] Jens Gregor and Thomas Benson. "Computational analysis and improvement of SIRT". In: *IEEE Transactions on Medical Imaging* 27.7 (2008), pp. 918–924.

[5] John E Dennis and Jorge J Moré. "A characterization of superlinear convergence and its application to quasi-Newton methods". In: *Mathematics of computation* 28.126 (1974), pp. 549–560.

[6] Jing Wang et al. "Penalized weighted least-squares approach to sinogram noise reduction and image reconstruction for low-dose X-ray computed tomography". In: *IEEE transactions on medical imaging* 25.10 (2006), pp. 1272–1283.

[7] Emil Y Sidky, Chien-Min Kao, and Xiaochuan Pan. "Accurate image reconstruction from few-views and limited-angle data in divergent-beam CT". In: *Journal of X-ray Science and Technology* 14.2 (2006), pp. 119–139.

[8] Lothar Reichel and Qiang Ye. "Simple square smoothing regularization operators". In: *Electronic Transactions on Numerical Analysis* 33 (2009), p. 63.

[9] Jorge Nocedal. "Updating quasi-Newton matrices with limited storage". In: *Mathematics of computation* 35.151 (1980), pp. 773–782.

[10] Richard H Byrd, Jorge Nocedal, and Robert B Schnabel. "Representations of quasi-Newton matrices and their use in limited memory methods". In: *Mathematical Programming* 63.1-3 (1994), pp. 129–156.

[11] Larry Nazareth. "A relationship between the BFGS and conjugate gradient algorithms and its implications for new algorithms". In: *SIAM Journal on Numerical Analysis* 16.5 (1979), pp. 794–800.

[12] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[13] Yu-Jung Tsai et al. "Fast Quasi-Newton Algorithms for Penalized Reconstruction in Emission Tomography and Further Improvements via Preconditioning". In: *IEEE transactions on medical imaging* 37.4 (2018), pp. 1000–1010.

[14] Per Christian Hansen. "The L-curve and its use in the numerical treatment of inverse problems". In: (1999).

[15] Julianne Chung, Malena I Español, and Tuan Nguyen. "Optimal regularization parameters for general-form Tikhonov regularization". In: *arXiv preprint arXiv:1407.1911* (2014).

[16] Michael F Hutchinson. "A stochastic estimator of the trace of the influence matrix for laplacian smoothing splines". In: *Communications in Statistics-Simulation and Computation* 19.2 (1990), pp. 433–450.

[17] Yifei Lou et al. "A weighted difference of anisotropic and isotropic total variation model for image processing". In: *SIAM Journal on Imaging Sciences* 8.3 (2015), pp. 1798–1823.

[18] Leonid I Rudin, Stanley Osher, and Emad Fatemi. "Nonlinear total variation based noise removal algorithms". In: *Physica D: nonlinear phenomena* 60.1-4 (1992), pp. 259–268.

[19] Antonin Chambolle and Thomas Pock. "A first-order primal-dual algorithm for convex problems with applications to imaging". In: *Journal of mathematical imaging and vision* 40.1 (2011), pp. 120–145.

[20] Ernie Esser, Xiaoqun Zhang, and Tony F Chan. "A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science". In: *SIAM Journal on Imaging Sciences* 3.4 (2010), pp. 1015–1046.

[21] Stephen Boyd et al. "Distributed optimization and statistical learning via the alternating direction method of multipliers". In: *Foundations and Trends® in Machine learning* 3.1 (2011), pp. 1–122.

[22] Sathish Ramani and Jeffrey A Fessler. "A splitting-based iterative algorithm for accelerated statistical X-ray CT reconstruction". In: *IEEE transactions on medical imaging* 31.3 (2012), pp. 677–688.

[23] Michael Lustig, David Donoho, and John M Pauly. "Sparse MRI: The application of compressed sensing for rapid MR imaging". In: *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine* 58.6 (2007), pp. 1182–1195.

[24] Lawrence A Shepp and Benjamin F Logan. "The Fourier reconstruction of a head section". In: *IEEE Transactions on nuclear science* 21.3 (1974), pp. 21–43.

[25] DR White et al. "ICRU Report 44: Tissue Substitutes in Radiation Dosimetry and Measurement". In: *J ICRU* (1989).

[26] Wim van Aarle et al. "The ASTRA Toolbox: A platform for advanced algorithm development in electron tomography". In: *Ultramicroscopy* 157 (2015), pp. 35–47.

[27] Folkert Bleichrodt et al. "Easy implementation of advanced tomography algorithms using the ASTRA toolbox with Spot operators". In: *Numerical algorithms* 71.3 (2016), pp. 673–697.

[28] G Zhang et al. "Bowtie filtration for dedicated cone beam CT of the head and neck: a simulation study". In: *The British journal of radiology* 86.1028 (2013), p. 20130002.

[29] Xiaoli Yang et al. "TV-based conjugate gradient method and discrete L-curve for few-view CT reconstruction of X-ray in vivo data". In: *Optics Express* 23.5 (2015), pp. 5368–5387.

[30] Patrick J La Rivière. "Penalized-likelihood sinogram restoration for CT artifact correction". In: *Nuclear Science Symposium Conference Record, 2004 IEEE*. Vol. 5. IEEE. 2004, pp. 3303–3307.

[31] Yan Liu et al. "Adaptive-weighted total variation minimization for sparse data toward low-dose x-ray computed tomography image reconstruction". In: *Physics in Medicine & Biology* 57.23 (2012), p. 7923.

[32] Friedrich L Bauer. "Optimally scaled matrices". In: *Numerische Mathematik* 5.1 (1963), pp. 73–87.

[33] Abraham Van der Sluis. "Condition numbers and equilibration of matrices". In: *Numerische Mathematik* 14.1 (1969), pp. 14–23.

[34] Tobias Lindstrøm Jensen and Moritz Diehl. "An Approach for Analyzing the Global Rate of Convergence of Quasi-Newton and Truncated-Newton Methods". In: *Journal of Optimization Theory and Applications* 172.1 (2017), pp. 206–221.

[35] Martin Burger and Stanley Osher. "A guide to the TV zoo". In: *Level set and PDE based reconstruction methods in imaging*. Springer, 2013, pp. 1–70.

[36] Robert E Alvarez and Albert Macovski. "Energy-selective reconstructions in x-ray computerised tomography". In: *Physics in Medicine & Biology* 21.5 (1976), p. 733.

[37] M. J. Berger et al. *XCOM: Photon Cross Sections Database*. URL: `http://physics.nist.gov/xcom`.

[38] Thorsten M Buzug. *Computed tomography: from photon statistics to modern cone-beam CT*. Springer Science & Business Media, 2008.

[39] Yi Zhang et al. "Spectral CT Reconstruction with Image Sparsity and Spectral Mean." In: *IEEE Trans. Computational Imaging* 2.4 (2016), pp. 510–523.

[40] Kaare Brandt Petersen, Michael Syskind Pedersen, et al. "The matrix cookbook". In: *Technical University of Denmark* 7.15 (2008), p. 510.

[41] E Roessl and R Proksa. "K-edge imaging in x-ray computed tomography using multi-bin photon counting detectors". In: *Physics in Medicine & Biology* 52.15 (2007), p. 4679.

[42] Carsten Oliver Schirra et al. "Statistical reconstruction of material decomposed data in spectral CT." In: *IEEE Trans. Med. Imaging* 32.7 (2013), pp. 1249–1257.

[43] Johann Niesen and Alex Sawatzky. "Performance of primal-dual algorithms for multi-channel image reconstruction in spectral X-ray CT". In: *Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2014 IEEE*. IEEE. 2014, pp. 1–7.

[44] William H Press et al. *Numerical recipes 3rd edition: The art of scientific computing*. Cambridge university press, 2007.

[45] Glen Cowan. *Statistical data analysis*. Oxford university press, 1998.

[46] J Punnoose et al. "spektr 3.0—A computational tool for x-ray spectrum modeling and analysis". In: *Medical physics* 43.8Part1 (2016), pp. 4711–4717.

[47] Taly Gilat Schmidt, Rina Foygel Barber, and Emil Y Sidky. "A Spectral CT method to directly estimate basis material maps from experimental photon-counting data". In: *IEEE transactions on medical imaging* 36.9 (2017), pp. 1808–1819.

[48] Emil Y Sidky et al. "Three material decomposition for spectral computed tomography enabled by block-diagonal step-preconditioning". In: *arXiv preprint arXiv:1801.06263* (2018).

[49] Kristian Bredies, Karl Kunisch, and Thomas Pock. "Total generalized variation". In: *SIAM Journal on Imaging Sciences* 3.3 (2010), pp. 492–526.

[50] Se Young Chun, Yuni K Dewaraja, and Jeffrey A Fessler. "Alternating direction method of multiplier for tomography with nonlocal regularizers". In: *IEEE transactions on medical imaging* 33.10 (2014), pp. 1960–1968.

[51] Michael Elad and Michal Aharon. "Image denoising via sparse and redundant representations over learned dictionaries". In: *IEEE Transactions on Image processing* 15.12 (2006), pp. 3736–3745.

[52] Lin Fu et al. "Comparison between pre-log and post-log statistical models in ultra-low-dose CT reconstruction". In: *IEEE transactions on medical imaging* 36.3 (2017), pp. 707–720.