Precise and Automated Tomographic Reconstruction with a Limited Number of Projections

Zur Erlangung des akademischen Grades eines
DOKTOR-INGENIEURS
an der Fakultät für
Elektrotechnik und Informationstechnik
des Karlsruher Instituts für Technologie (KIT)
genehmigte Dissertation
von

M.Eng. Xiaoli Yang

geb. in Shandong Province, China

Hauptreferent: Prof. Dr. rer. nat. Marc Weber
Korreferent: Prof. Dr. rer. nat. Olaf Dössel

Tag der mündlichen Prüfung: 12.01.2016

Ich versichere wahrheitsgemäß, die Dissertation bis auf die dort angegebene Hilfe selbständig angefertigt, alle benutzten Hilfsmittel vollständig und genau angegeben und alles kenntlich gemacht zu haben, was aus Arbeiten anderer und eigenen Veröffentlichungen unverändert oder mit Änderungen entnommen wurde.

Karlsruhe, den

_____

Xiaoli Yang

Abstract

X-ray computed tomography (CT) is a popular, non-invasive technique capable of producing high spatial resolution images. It is generally utilized to provide structural information of objects to examine. In multiple applications, such as biology, material, and medical science, more and more attention has been paid to tomographic imaging with a limited amount of projections. In this thesis, the tomographic imaging with a limited amount of projections is important in adapting a rapid imaging process or limiting the X-ray radiation exposure to a low level. However, fewer projections normally imply poorer quality of image reconstruction with artifacts by the traditional filtered back-projection (FBP) method, hampering the image analysis and interpretation. A new reconstruction method other than the FBP method is demanded for reducing the projections from over a thousand to around a hundred. Guided by compressive sampling (CS) theory, iterative reconstruction technique based on sparsity helps to overcome these difficulties; however the iterative reconstruction has a number of drawbacks that are tackled in this thesis to allow an automated, fast and high quality reconstruction on a computer cluster.

The main component of the framework and contribution of this thesis is a parameter-optimized iterative reconstruction method (optimized-CGTV). It is proposed for tomographic reconstruction with limited projections subject to the minimization of the total variation (TV). This inverse problem is formulated in a Lagrangian multiplier fashion with the parameter optimization by applying a discrete L-curve obtained from the plot of the TV regularization versus data fidelity. The reconstruction is carried out using the corresponding developed conjugate gradient solver to minimize the TV-regularized formulation. The optimized-CGTV reconstruction method is incorporated into an automatic framework of parallel 3D reconstruction on a computer cluster to achieve a rapid reconstruction process.

The proposed method is evaluated with various datasets, including both simulated and synchrotron X-ray tomographic data (such as the imaging of grain weevil and frog embryo). The reconstructions of simulated and experimental data from around 100 projections separately demonstrate the high quality reconstruction by the optimized-CGTV method in both noiseless case and cases subject to different levels of noise. The optimized-CGTV reconstruction reduces strealine artifacts, which appear in the FBP reconstruction due to insufficient projection data, while well preserving main edges of homogeneous structures. With the parallel reconstruction framework in the computer cluster, a 3D volume reconstruction can be completed in the order of 10 minutes with a speedup factor over a hundred compared to the sequential implementation. This reconstruction performance is accepted as near real-time for offline analysis.

The method in this thesis has been applied to three tomographic data sources: X-ray tomographic imaging at the TOPO-TOMO bending-magnet beamline of ANKA, X-ray tomographic imaging at the undulator beamline 32-ID of APS, and electron tomographic imaging for nanostructured material analysis. In all these applications, the optimized-CGTV method can automatically produce high quality reconstruction from a limited number of projections. As a result, this work enables the building of fast tomography at ANKA, constrains the X-ray radiation dose to extend the scale of *in vivo* time-lapse series for developmental biology, and allows more precise analysis of nanostructured materials using electron tomography.

Zusammenfassung

Röntgen-Computertomographie (CT) ist eine etablierte, nichtinvasive Technik zur Erzeugung hochauflösender Volumenbilder. Sie wird üblicherweise dazu benutzt, Strukturinformationen untersuchter Objekte zu gewinnen. In Anwendungsbereichen wie der Biologie, der Medizin und den Materialwissenschaften wird zunehmend Augenmerk auf Tomographie mit wenigen Projektionen gelegt. Diese ist beispielsweise bei schnellen Aufnahmeverfahren oder speziellen Aufnahmegeometrien von Bedeutung, oder dann, wenn die Strahlungsdosis auf ein geringes Maß reduziert werden soll. Eine geringere Zahl an Projektion bedingt jedoch eine verschlechterte Bildqualität mit Rekonstruktionsartefakten, so dass es zu Problemen in der anschließenden Bildanalyse kommt. Dies legt den Bedarf an einer neuen Rekonstruktionsmethode nahe, die nicht den Einschränkungen der traditionell verwendeten gefilterten Rückprojektion (Filtered Back-Projection, FBP) unterliegt und Artefaktbildung vermeidet. Unter Zuhilfenahme der Compressive-Sensing-Theorie (CS-Theorie) können iterative Rekonstruktionsmethoden, die sich die Eigenschaft der Dünn-Besetztheit (Sparsity) zunutze machen, solche Rekonstruktionsprobleme verhindern. Die Anwendung solcher Methoden ist jedoch durch die Parameterabhängigkeiten, die üblicherweise empirische Parameterbestimmung, den manuellen Implementierungsaufwand und die hohe Rechenzeitanforderungen schwierig.

Um dieser Herausforderung zu begegnen, wird in der vorliegenden Arbeit eine parameteroptimierte iterative Rekonstruktionsmethode für die Computertomographie aufgrund weniger Projektionen vorgestellt. Wichtigstes Optimierungskriterium bei der Rekonstruktion eines präzisen 3D-Bildes ist dabei eine Regularisierung der totalen Variation (TV) im Sinne der CS-Theorie. Die iterative Rekonstruktion wird mit Hilfe der Methode der Lagrange-Multiplikatoren formuliert, wobei die Parameteroptimierung in der Anwendung der diskreten L-Kurventheorie besteht, bei der die Datentreue und die TV-Regularisierung gegeneinander abgewogen werden. Diese Rekonstruktionsmethode wird in ein eigens entwickeltes, automatisches Framework eingebettet, das zur schnellen, parallelen 3D-Rekonstruktion in einem Rechencluster dient.

Das automatische Framework wird anhand verschiedener Datensätze evaluiert, die sich sowohl aus simulierten als auch aus echten Synchrotron-Röntgentomographiedaten zusammensetzen. Für die Echtdaten wurden unter anderem ein Kornkäfer als auch ein Froschembryo untersucht. Die Rekonstruktionen der simulierten und der echten Daten mit einer begrenzten Zahl von Projektionen demonstrieren die hohe Bildqualität der entwickelten Methode sowohl im rauschfreien als auch im Fall unterschiedlich starken Rauschens. Mit dem entwickelten parallelen Rekonstruktionsframework ist es möglich, 3D-Volumenrekonstruktionen mit einem hohen Beschleunigungsfakor gegenüber der sequentiellen

Implementierung durchzuführen, womit nun Nah-Echtzeit-Rekonstruktionen für Offline-Analyse möglich sind.

Die vorgestellte Methode wird für drei verschiedene tomographische Datenquellen eingesetzt: Röntgentomographie der TOPO-TOMO-Beamline (Bending-Magnet-Technik) bei ANKA, Röntgentomographie in der Undulator-Beamline 32-D bei APS und Elektronentomographie mit Nanometer-Auflösung. In alle Anwendungen lieferte die entwickelte Methode präzise Rekonstruktionsergebnisse bei eine geringen Anzahl von Projektionen. Die geringe Zahl an Projektionen geht mit geringerer Strahlenbelastung bei gleichbleibender Bildqualität einher und ermöglicht beispielsweise in der Entwicklungsbiologie, eine maßgebliche Reduktion der Expositionszeit bei In-Vivo-Messserien. Im Fall der Elektronentomographie ist nun eine genauere Analyse nanostrukturierter Materialien möglich.

Contents

1. Introduction

X-ray computed tomography has demonstrated dramatic impact on various fields due to its superior imaging quality and has become one of the most commonly used imaging techniques over the last several decades. The corresponding reconstruction approach has also been greatly improving to overcome the newly emerging difficulties and satisfy the changing requirements. This chapter will introduce the scientific study using X-ray computed tomography in terms of the application background, motivations for sparse-view reconstruction, the state of the art, challenges, and proposed solution.

1.1. Background

The imaging technology based on X-ray radiation was developed since the year of 1895, when German physicist Wilhelm Conrad Röntgen discovered the electromagnetic radiation of X-rays, also known as Röntgen rays. This enables a noninvasive insight into the human body for the first time and thus significantly contributes to the progress of modern medical imaging [1]. The mathematical foundation of computed tomography (CT) was provided by the Radon transform in 1917 [2]. In the late 1960s and early 1970s, the American physicist Allan MacLeod Cormack and the English electrical engineer Godfrey Newbold Hounsfield developed the X-ray CT technology [3] to produce a large series of two-dimensional (2D) radiographic images of a scanned object, from which a three-dimensional (3D) image is reconstructed, allowing people to see the inner structures of the object without destroying it.

As a 3D imaging technique, computed tomography has significant advantages over the traditional 2D radiography [4]. Firstly, it completely eliminates the superimposition of structures in the final image. Secondly, CT has inherent high-contrast resolution, so that the variation in the physical density of different tissues can be distinguished effectively. Besides, the final reconstructed 3D image can be reformulated in mutiplanar way that the object can be viewed in the axial, coronal or sagittal planes depending on the specific requirements.

The most common application of X-ray CT is medial imaging, in which the cross-sectional reconstruction images are used for diagnostic and therapeutic purposes in various medical disciplines. Nevertheless, the X-ray CT technology is also importantly applied in

other fields, such as biology, life science, material science, industrial scanning, and so on. As the X-ray CT imaging technique is good in providing structural information with high spatial resolution, it is often fused with other imaging modalities, like ultrasonography (US) [5], magnetic resonance imaging (MRI) [6], positron emission tomography (PET) [7], single photon emission computed tomography (SPECT) [8], and optical imaging (OI) [9]. The combination with one of these techniques allows to locate the position of diseases, defects, or other type of targets. Thus the X-ray CT technique plays an important role in wide application fields.

However, 3D computed tomography is only able to image the stationary structure of the object. In order to study and characterize the object motion or development, time-resolved volumetric CT, also known as four-dimensional (4D) *in vivo* tomographic imaging, is used and becoming more and more popular in scientific study [10]. The time information required for motion or development characterization in 4D imaging can be obtained by fast imaging at discrete time points. At each time point, a number of snapshot projections are acquired for reconstructing a 3D image without motion artifacts. A movie of all these time-resolved 3D images presents the dynamic characters to track target motion or variations of critical tissue structures.

A representative example of 4D *in vivo* tomographic imaging is the investigation made by Thomas van de Kamp [11] regarding a moving screw-and-nut-type hip joint in the insect *Sitophilus granarius* (grain weevil) [12] at the synchrotron light source ANKA [13] located at Karlsruhe Institute of Technology (KIT). Fig. 1.1 displays the scanning electron microscope (SEM) photographs of the joint, including the right hind hip in 1.1a showing the notch (n) and inner thread (i) and the corresponding trochanter part in 1.1b showing the external thread (e) and posterior thorn (t) in the weevil.

The structure and function of the screw-and-nut system are revealed through the 3D reconstructions created from the *in vivo* synchrotron-based X-ray tomography. The experimental setup of the synchrotron tomography beamline is shown in Fig. 1.2a including the X-ray source (here: bending magnet), tomographic stage, sample, and detector system. The real sample fixed on the stage is displayed in the photograph in Fig. 1.2b. By rotating the sample many tomographic projections in different views are recorded by the detect system, one of which is shown in Fig. 1.2c. From these projections the scientist will apply a reconstruction method to achieve the 3D structure of the weevil. The 3D illustration of the screw joint of the weevil in Fig. 1.2d is finally obtained by segmenting the reconstruction result to distinguish the inner organs and tissues, such as muscles (m1, m2, m3a, and m3b), coxa, trochanter, and femur. A series of 3D reconstructions from the *in vivo* tomographic data in different time lapses will illustrate the dynamic process of the screw joint.

This investigation applies the X-ray phase-contrast computed microtomography (XPCμT) together with absorption effect to enhance the contrast of biological tissues which have low difference in X-ray attenuation coefficients. It is able to produce projections with

(a)   (b)
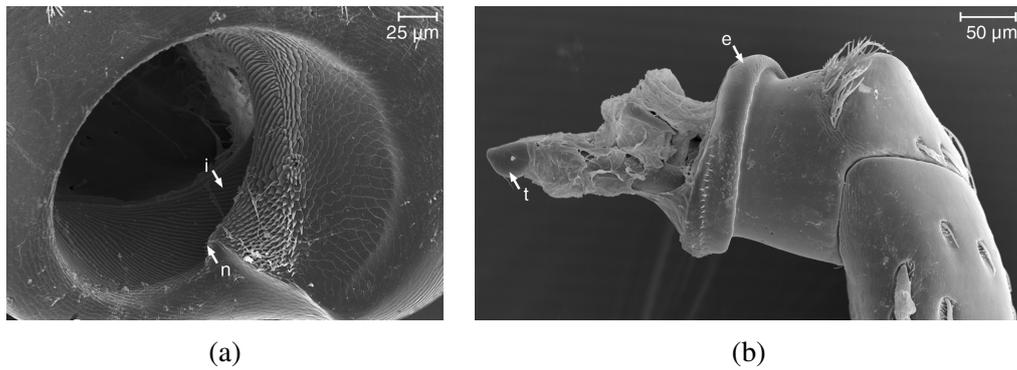
Fig. 1.1. Scanning electron microscope (SEM) photographs of a screw-and-nut-type joint: (a) is the right hind hip showing the notch (n) and inner thread (i); (b) is the corresponding trochanter part showing the external thread (e) and posterior thorn (t).

(a)   (b)

(c)   (d)
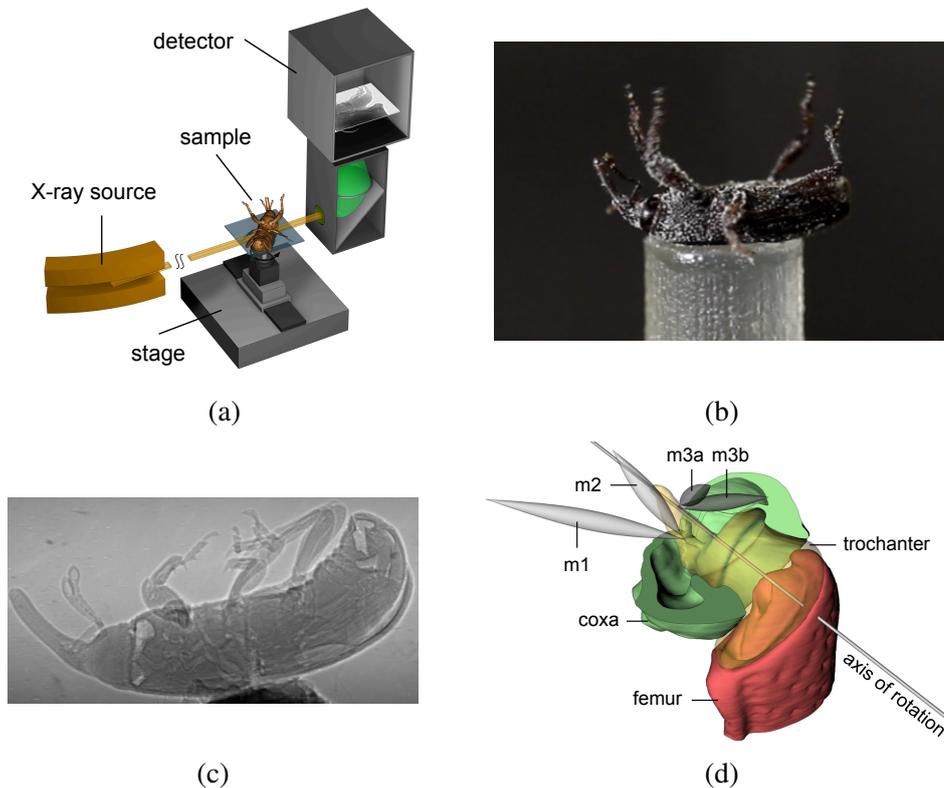
Fig. 1.2. Investigation of a moving screw-and-nut-type hip joint using X-ray tomography: (a) experimental setup of the synchrotron tomography beamline; (b) photograph showing the real weevil on the stage; (c) single radiographic projection recorded by the detector system; (d) 3D illustration of the screw joint of the weevil (m1, m2, m3a, and m3b are muscles), created from the 3D reconstruction of projections in (c).
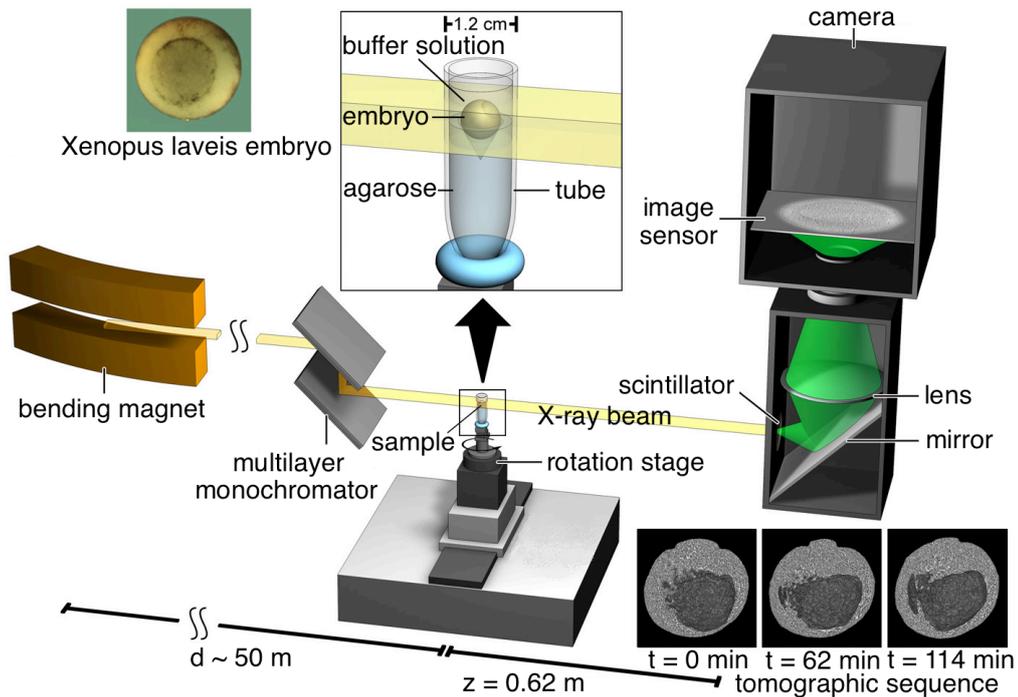
Fig. 1.3. Experimental setup for X-ray phase-contrast computed microtomography (taken from the article [14] by Julian Mossmann). X-ray beam generated from a bending magnet propagate over a large distance (~ 50 m) to impinge on the sample (living *Xenopus laevis* embryo) mounted on a rotation stage for tomographic projections acquired by image sensor and camera.

sub-micron spatial resolutions thanks to the high degree of spatio-temporal coherence and high flux density in state-of-the-art synchrotron radiation. Additionally the impressive advances in detector technology makes exposure time below one millisecond per projection feasible. Thus the investigation of a moving insect can be practically performed for the elucidation of rapid dynamics. In particular, *in vivo* elucidations of internal processes in short time lapses benefit from the highest possible volume-acquisition rates.

The second instance for the 4D in vivo tomographic imaging is the study of embryonic development of vertebrate model organism *Xenopus laevis* (African clawed frog) made by Julian Moosmann [14]. It aims to observe the behavior of cells and important processes during gastrulation by *in vivo* imaging the changes of the embryonic structure via 3D time-lapse series [14, 15]. The fourth dimension beyond the 3D space is time, in which the embryonic development is recorded. The investigation of *Xenopus* is important for the complete understanding of cloning and developmental processes [14].

Since the frog's embryo is optically opaque, of small size in micron scale, and lack of considerably absorptive structures (e.g. bone), it is difficult to find a high-resolution imag-

ing method for the observation of cell and tissue dynamics, including motion and shape changes. This problem can be solved in phase contrast images which are obtained by the technique of X-ray phase-contrast computed microtomography resolving cell boundaries and tissue structures [16, 17]. The experimental setup for X-ray phase-contrast microtomography is demonstrated in Fig. 1.3 in detail. It is quite similar with the setup in Fig. 1.2a but has larger distance from the X-ray source to the sample for generating phase-contrast images. The upper left and lower right corner in Fig. 1.3 respectively displays a photograph of the *Xenopus laevis* embryo and a sequence of images created from tomographic reconstruction in different time lapses. Note that a step of phase retrieval has to be performed to restore the projections just before tomographic reconstruction. This experiment was carried out at the undulator imaging beamline 32-ID of the Advanced Photon Source (APS) in Argonne, Illinois, America [15].

There are surely many more interesting 3D or 4D tomographic imaging applications other than the two above mentioned instances. For all these applications, the 3D reconstruction in a short time-lapse is an issue of key importance. It recovers the 3D distribution of attenuation coefficients from a large amount of 2D tomographic projections, thus enabling the observation of inner structures of the object in three dimensions. Scientific researchers generally pay more attention to two terms of the CT reconstruction: accuracy and speed. On the one hand, a high quality reconstruction is demanded for correct and precise analysis of obtained object data, such as the location, segmentation, and feature extraction of a particular target of interest. This can be achieved by building up an accurate system model for the forward imaging process to ensure data fidelity [18, 19] and also by applying advanced prior knowledge-based algorithms to confine artifacts [20, 21]. On the other hand, a high-speed reconstruction is important for the practical purpose that scientific researchers do not wait too long for data reconstruction and analysis processes. This greatly raises the efficiency and decrease the intensity of labor. The reconstruction is often accelerated in two ways, one of which utilizes advanced computing hardware and structures, such as muticore computers [22], computer cluster [23], and graphics processor units (GPUs) [24] for parallel and fast computing. Another way focuses on the acceleration in terms of reconstruction algorithm itself by replacing the original reconstruction problem with a new one [25] or applying an iteration strategy with fast convergence [26]. As the forward operation implemented many times in the reconstruction algorithm and involving large matrices multiplication, it takes a large part of the whole reconstruction time. Some researchers also resort to fast implementation of the forward model for the acceleration of reconstruction algorithm [27, 28, 29]. However, an accurate reconstruction often involves more complicated operations in the reconstruction algorithm which are normally more expensive in computation. Therefore a practical compromise must be made between accuracy and speed taking into account the computation power available.

The reconstruction methods were greatly developed with arising of new requirements and challenges. Until now various reconstruction methods have been presented, which can be classified into two categories: analytical and iterative methods. The analytical re-

construction methods such as the standard filtered back-projection method (FBP) [30] are commonly used in commercial CT systems and also scientific research using tomographic technology. Another type of reconstruction methods, iterative methods such as algebraic reconstruction technique (ART) [30, 31, 32], search for a better reconstructed solution in a high dimensional domain in each iteration until it converges to an optimal one. Since these two categories of reconstruction methods mathematically apply very different principles in performing reconstruction, they show different features and capabilities in reconstructing the original object from various projections. The FBP reconstruction is a simple and fast implementation, insensitive to noise in projections in the case of sufficient views available, but it generates conspicuous artifacts for insufficient data. By contrast, the iterative method is able to obtain a better reconstruction if the projections are not distributed uniformly in angle or if the projections are sparse or missing in certain orientations. But it is at the cost of a longer computation time, and high noise in projections can cause a large data inconsistency leading to relatively inaccurate reconstruction result for iterative method. Generally speaking, the iterative reconstruction method draws more and more attention at present. Various of algorithms have been developed in this direction to support reconstructions with different requirements. Even though these sophisticated reconstruction methods have existed, the tomographic reconstruction is still challenging in many situations in which the tomographic imaging has to be performed with sparse-view projections, such as the two 4D *in vivo* tomographic imaging instances discussed above.

1.2. Motivations for sparse-view reconstruction

ANKA is a synchrotron light source [33] founded since 2003 at the north campus of KIT (Karlsruhe Institute of Technology), Karlsruhe, Germany. It provides light from hard X-rays to far-infrared and experimental platforms for various researches and technologies. The scientific topic in concern here is the fast tomography beamline [13] at ANKA using X-rays for 4D *in vivo* imaging as shown in Fig. 1.2. It allows high-resolution imaging ($<$ 1 μm) of moving objects in a short time lapse ($<$ 10 ms) per tomographic scan (tomogram, including projections in [0°, 180°]), thus avoiding the artifacts and uncertainties introduced by object motion. The high-resolution and high-speed properties require a high-throughput data acquisition (DAQ) system which is developed by the project UFO (ultra-fast X-ray imaging) [34]. It benefits from the integrated high-speed camera which can achieve several thousand frames per second and enables the continuous data acquisition at full speed. However, even though the advanced high-speed camera is used, the demanded short time lapse for each tomographic scan still limits the number of projections to be much fewer in the range from several dozen to several hundred instead of over thousand in a tomogram. Thus in this instance, we are confronted with the problem of insufficient projections for tomographic reconstruction.

In the second investigation of the embryonic development as shown in Fig 1.3, the greatest limitation is the X-ray radiation dose absorbed by the embryo per tomogram. The dose restricts the length of the time-lapse series to around two hours of development. Even though it is yet sufficient for tracing many important developmental events, dose reduction is able to extend this time scale significantly. Two strategies of data acquisition for reducing the radiation dose can be explored. One strategy is to simply reduce the milliampere-seconds (mAs) or the kilovoltage-peak (kVp) values per view in the data acquisition process [35, 36, 37]. This approach, however, results in insufficient X-ray photons received by detectors, increasing the noise level in the projections. This will degrade the quality of tomographic reconstruction. Another way to limit the total radiation dose is to reduce the number of tomographic projections per rotation around the object [20, 38]. It can accelerate the data acquisition process and thus reduce the exposure time of the object to X-ray radiation allowing extended observation of developing embryo. Thus again in this case we are confronted with the same problem as in the former one.

In the tomographic imaging cases mentioned above, a low number of tomographic projections are actually available due to different requirements. Thus the challenge for 3D tomographic reconstruction is to generate relevant information from limited input data. This work will consider only the parallel-beam CT used in our both tomographic imaging cases. But the reconstruction principles and strategies used in this work are also applicable to the fan- or cone-beam CT systems which are not be considered here. The main idea behind parallel-beam CT is that the projections at different angles are generated by shooting X-rays through an object of interest. According to the Fourier-slice theorem [30], the projections of a 2D cross-section can be represented as radial lines in Fourier domain, from which one can obtain the Fourier transform of the cross-section image. Namely, as shown in Fig. 1.4 each radial line denotes a projection in Fourier domain. Using inverse Fourier transform [39] it is possible to retrieve a good approximation of the object from its equally spaced and sufficient projections in Fig. 1.4. The density of theses radial lines in Fourier domain depends on the number of projections taken in different angles. Figure 1.5 displays two different sampling patterns, where the green and blue points denote the central positions of X-ray source and detector respectively. Figure 1.5a is the sampling pattern for an ordinary CT system which includes the projection data in over a thousand angles, and 1.5b gives the sparse-view pattern for CT system which samples the object in much fewer angles (around a hundred). This thesis will focus on the sparse-view CT system pursued in the investigations of Fig. 1.2 and 1.3.

The FBP is a commonly used reconstruction method based on a combination of the Fourier-slice theorem and the Nyquist-Shannon sampling theory. A high quality FBP reconstruction generally requires the number of equiangular-spaced, tomographic projections given as

$$P_{\text{FBP}} \approx \frac{\pi}{2} K \, . \tag{1.1}$$
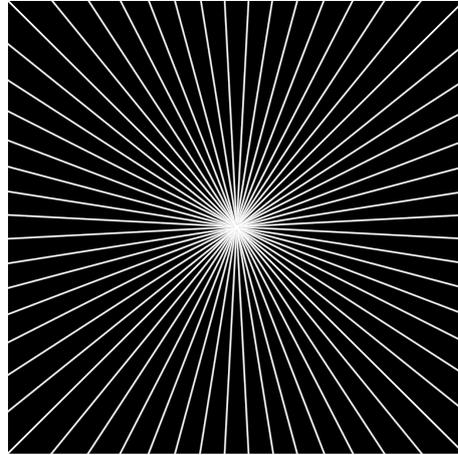
Fig. 1.4. Representation of projections in different angles in Fourier domain, from which one can obtain the Fourier transform of the cross-section image and recover the original object.

(a) ordinary CT (b) sparse-view CT
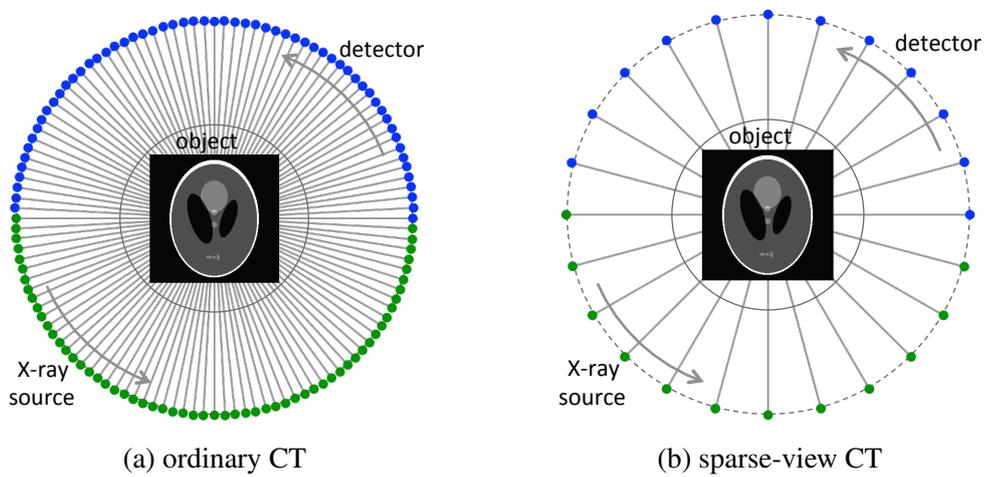
Fig. 1.5. Two different sampling patterns: (a) dense sampling for ordinary CT system generating 1000-2000 projections and (b) sparse-view sampling CT system generating ~ 100 projections.

(a)	(b)	(c)
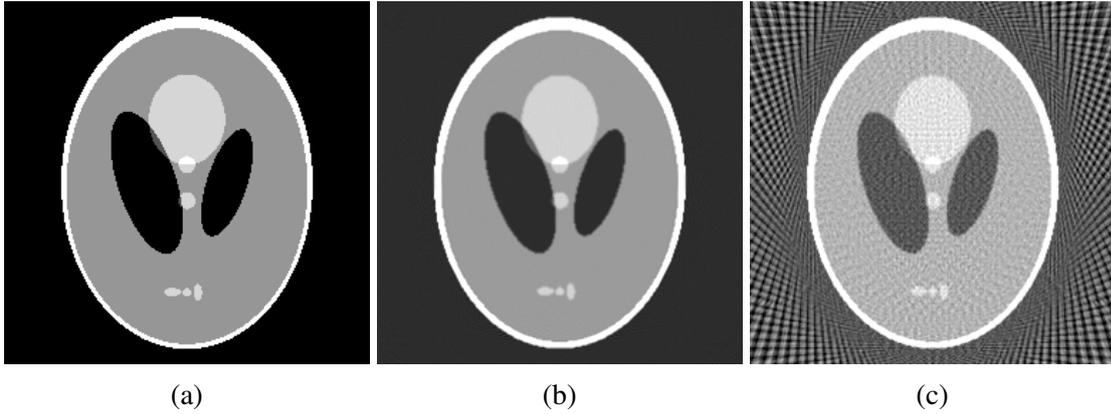
Fig. 1.6. FBP reconstructions of Shepp-Logan phantom with different number of projections: (a) shows the original phantom; (b) is reconstructed from 403 projections according to Eq. 1.1, providing a high quality reconstruction; however, a reconstruction, performed using only 60 projections in (c), shows serious streakline artifacts.

Here $K \equiv L/\Delta x$ with L representing the linear dimension of the (quadratic) field of view. For more, see Sec. 3.2. Sparsely sampled projection data result in a degraded quality of FBP reconstruction [40]. This can be seen from the reconstructions in Fig. 1.6 for a Shepp-Logan phantom (256×256), which is a model of a human head created in [41] and often serves as a standard test image for image reconstruction algorithms. According to Eq. 1.1 where $K = 256$, around 403 projections are required for a satisfactory reconstruction which is shown in Fig. 1.6b. A reconstruction from limited data of only 60-view projections is shown in Fig. 1.6c, where serious streakline artifacts are caused by the FBP method. The reason for these streakline artifacts are the missing angular information in the Fourier domain [42]. Condering the representation of projections in Fourier domain in Fig. 1.4 in the case of sparse-view CT, it is not possible to precisely recover the lost information between every two adjacent radial lines by using interpolation method. The corresponding artifacts becomes more serious when it goes far away from the center of the image. It is apparently difficult to continue data analysis with the sparse-view reconstruction. In summary, the FBP is a commonly used, rapid reconstruction technique, but lacks of ability to process the situation of sparse-view reconstruction.

1.3. State of the art

The drawback of FBP reconstruction method encourages scientific researchers to pursue ways of obtaining high quality reconstruction images utilizing as fewest projections as possible. Fortunately, the objects of interest tend to benefit from the property of being sparse. In other words, the image of an object often consists of constant values in large

(a) non-sparse representation   (b) sparse gradient image
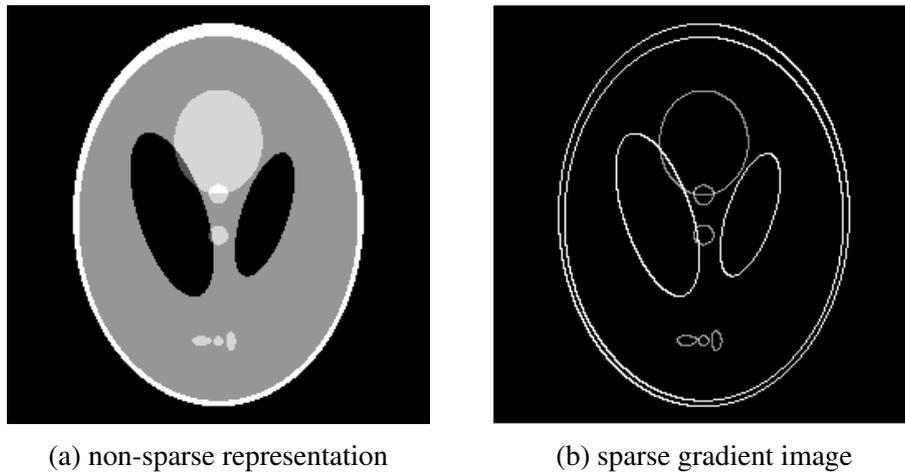
Fig. 1.7. Sparse representation of Shepp-Logan phantom in a transformed domain: (a) is the original Shepp-Logan phantom (256×256); (b) shows the distribution of the phantom after the discrete gradient transform (DGT).

regions. It means there exist large redundancies in the representation of the object. It is possible to find a more sparse representation, which suggests a possible dimensionality reduction thus allowing Fourier sampling in the angular direction at lower rate.

For example, pursuing this idea, a base change called discrete gradient transform (DGT) can lead to a sparse expansion of a certain 2D or 3D image. The two different representations for the famous Shepp-Logan phantom are shown in Fig. 1.7. In the original phantom (256×256) in Fig. 1.7a, 27409 pixels out of the total 65536 are non-zero, which means that the phantom is not sparsely represented in the spatial domain. In contrast, the gradient image of the phantom, transformed by DGT, shows high sparsity in Fig. 1.7b. In this image only 2183 non-zero elements exist, taking 3.3% of the total pixels, which definitely demonstrates a sparse representation.

The theory behind this idea is called compressive sampling (CS), which claims that, if the object has a sparse representation in certain domain, the original object can be precisely restored from a set of measurements with much lower sampling rate required by Nyquist-Shannon theory. This theory gives definite conditions on the number of required measurements for a sparsely represented object [42] under which the object can be recovered within a prescribed error. In image reconstruction, an alternative approach instead of FBP method, iterative reconstruction (IR) technique [30, 31, 32], is able to accommodate such prior knowledge as the existence of a sparse representation, and thus one may hope that a limited set of noisy data suffices in reconstructing such an object with a good accuracy.

The iterative reconstruction technique starts from the creation of a linear problem from a forward model representing the imaging process. The basic linear problem can be mod-

eled by the simple equation system:

$$\mathbf{A}\mathbf{x} = \mathbf{p}\,, \tag{1.2}$$

where $\mathbf{A}$ is a system matrix derived from the forward model; $\mathbf{x}$ is an object image indicating the distribution of the attenuation index; and $\mathbf{p}$ denotes a vector embodying the set of linear measurements of $\mathbf{x}$ (here are projections acquired during the tomographic imaging process). The dimensionalities of $\mathbf{A}$, $\mathbf{x}$, and $\mathbf{p}$ depend on the size of discretized object image and the number of measured projections. The reconstruction method now performs the inverse operation of recovering $\mathbf{x}$ in terms of $\mathbf{A}$ and $\mathbf{p}$. However, the inverse problem of Eq. 1.2 is very ill-posed due to two points. First, as we are considering the sparse sampling pattern shown in Fig. 1.5b, the resulting number of equations in Eq. 1.2 lies essentially below the bound of Eq. 1.1 and also the number of unknown elements in $\mathbf{x}$. Secondly, there always exist errors in the real measurements $\mathbf{p}$, that is $\mathbf{A}\mathbf{x} = \mathbf{p}_{\text{ture}} + \varepsilon$. The measured data are not consistent with the model. The property of being ill-posed leads to no exact solution to Eq. 1.2. Thus the determination of an appropriate solution demands some sort of a prior knowledge and/or the application of deep physical principles. The process of constructing a unique solution subject to such additional conditions is called regularization.

To pursue the theory based on sparse approximation, a sequence of base changes, transforming the original (spatial domain) into an increasingly sparse set of representations, is required for constructing an optimal representation of the object. Suppose that the base changes are denoted by Φ and $\mathbf{x}'$ is the sparse representation of $\mathbf{x}$ in the domain Φ. Thus we have $\mathbf{x}' = \Phi\mathbf{x}$ and $\mathbf{x} = \Phi^{-1}\mathbf{x}'$. The inverse problem of Eq. 1.2 can be imposed as an optimization problem using least square principle and regularization like:

$$\min_{\mathbf{x}}\{\|\mathbf{A}\Phi^{-1}\mathbf{x}' - \mathbf{p}\|_2^2 + \lambda\|\mathbf{x}'\|_1\}\,. \tag{1.3}$$

Here the l_2-norm of error $\mathbf{A}\Phi^{-1}\mathbf{x}' - \mathbf{p}$ and the l_1-norm of $\mathbf{x}'$ in the regularization are minimized considering the data consistence and sparse representation respectively. The l_1-norm of a vector $\mathbf{b}$, $\|\mathbf{b}\|_1$, is with $\|\mathbf{b}\|_1 \equiv \sum_i |b_i|$ and the l_2-norm $\|\mathbf{b}\|_2$ is with $\|\mathbf{b}\|_2 \equiv (\sum_i b_i^2)^{1/2}$. This optimization aims to minimize the l_1-norm function while ensuring the solution to be consistent with measurements $\mathbf{p}$. Parameter λ controls the trade-off between data consistence and sparse regularization. One often writes formula (1.3) with respect to $\mathbf{x}$ instead of $\mathbf{x}'$ such as:

$$\min_{\mathbf{x}}\{\|\mathbf{A}\mathbf{x} - \mathbf{p}\|_2^2 + \lambda\|\Phi\mathbf{x}\|_1\}\,. \tag{1.4}$$

This minimization problem conforms to the CS theory and the transform base Φ can be any domain in which the original image is expressed sparsely. Here only the typical discrete gradient transform (DGT) shown in Fig. 1.7b is presented. It transforms the original image to its gradient image by integrating its gradients in horizontal and vertical direc-

tions. The summation of this gradient image over all coordinates forms the so-called total variation (TV) [43], that is also the l_1-norm of the gradient image vector. Thus the regularization based on DGT for Φ in Eq. 1.4 is often written as TV-norm like:

$$\min_{\mathbf{x}}\{\|\mathbf{Ax} - \mathbf{p}\|_2^2 + \lambda\|\mathbf{x}\|_{\mathrm{TV}}\}\,, \tag{1.5}$$

where $\|\mathbf{x}\|_{\mathrm{TV}} = \|\mathrm{DGT}(\mathbf{x})\|_1$. For more expressions, see Sec. 4.2. Note that the optimization in (1.5) is sometimes written in the a constrained form as

$$\min_{\mathbf{x}} \|\mathbf{x}\|_{\mathrm{TV}}, \quad \text{subject to} \quad \|\mathbf{Ax} - \mathbf{p}\|_2^2 \leq \epsilon\,. \tag{1.6}$$

It aims to minimize the total variation function while enforcing the optimal solution to be consistent with measurements $\mathbf{p}$. ϵ is a tolerance parameter required to obtain an approximate solution. The formulation in (1.5) can be literally regarded as a variant of (1.6) by introducing a Lagrangian multiplier [44].

The reconstruction method now searches an approximate solution to this optimization problem given above. Literally the algorithms developed for solving this problem use many iterations, in each of which they determine a better solution close to the optimal one, thus they are so called iterative methods. To solve the optimization problem in Eq. 1.5, one can resort to the generic optimization methods [45], such as the gradient descend method (also known as steepest descend method), nonlinear conjugate gradient method, and Newton's methods. One has to take into account the convergence speed, existence of necessary information, and computational demands. Specifically, the gradient descend method and nonlinear gradient method require only simple first-order information of the objective function in the optimization problem, but show lower convergence speed than Newton's methods. The Newton's methods can achieve fast convergence speed by using higher-order information which, however, may not exist or require very expensive computation for a large scale problem such as the tomographic reconstruction problem. In the literature, researchers are more dedicated to developing advanced algorithms which utilize only first-order information of the objective function while reaching an approximate solution in a fast speed even for a large scale problem.

Inspired by the CS theory, iterative reconstruction method with TV regularization has been studied for both few-view and limited angle problems [20]. In 2006, Emil Y. Sidky et al. introduce the concept of TV into clinical application and developed the algorithm TV-POCS for CT reconstruction [20]. It solves the constrained program (1.6) via an iteration of two steps: the minimization of the TV-norm performed by the gradient descent method and the constraint incorporated by projection on convex sets (POCS). TV-POCS algorithm was latter improved towards better convergence and robustness against artifacts by using the steepest descent method with an adaptive step-size (called ADS-POCS) [38].

Note that TV minimization is also applied to image denoising and restoration. In the general context, many solvers have already been developed for the TV-based, large-scale minimization problems (1.5) or (1.6) with fast convergence, such as TwIST [46] by Bioucas-Dias J. M. et al. in 2007, FISTA [26] by Beck A. et al. in 2009, NESTA [47] by Becker S. et al. in 2011, and UPN [48] by Jensen T. L. et al. in 2012. TwIST is a fast implementation of iterative shrinkage/thresholding (IST) algorithm especially when the inverse problem is ill-posed. Its combination with the TV minimization leads to a fast convergence at the first step. FISTA is another algorithm that can speed up the performance of IST method by computing the next iteration based not only on the previous one but on two or more previous iterations. The NESTA algorithm can also be easily extended for TV minimization by modifying slightly the smooth approximation of the objective function. UPN is an efficient method by further applying accelerated scheme of the NESTA algorithm. All of these algorithms are based on the iterative shrinkage/thresholding method but achieve faster convergence speed for iterations with only first-order gradient information required.

Moreover, another set of algorithms employs the splitting idea [49] in developing an alternating minimization approach for image recovery with TV regularization, such as TVAL3 [50, 51] presented by Li C. et al. in 2009, RecPF [52] by Yang J. et al. in 2010, and ADMM [53, 25] by Goldstein T. in 2012. They utilize techniques like the split Bregman algorithm [54], the augmented Lagrangian method or the alternating direction method and gain fast convergence speed. These techniques are not completely independent but have similar idea for the iterations. They employ a new variable to substitute the expression in regularization. As a result, the solution to the minimization problem can be achieved by alternately minimizing two new problems, in each of which one variable is updated while fixing the other one. The interrelations of these techniques were also pointed out in [55] and [56].

1.4. Challenges

From the perspective of a biologist, tomographic imaging is the method enabling the observation and analysis of biological functionality and development. Data acquisition and tomographic reconstruction are two key steps in this process, where advanced computer techniques or computer-aided methods can be used to improve the efficiency of data processing tasks like the automation of data acquisition, storage, transmission, reconstruction, and intelligent analysis processes. The 4D tomographic imaging shown in Fig. 1.2 and 1.3 are experiments involving a high-speed and high throughput procedure to generate large amount of data in time lapses. The acquisition, management, and processing of this kind of data requires a smart experimental station and intelligent intensive data services. A new smart experimental station is under construction as the main goal of the project
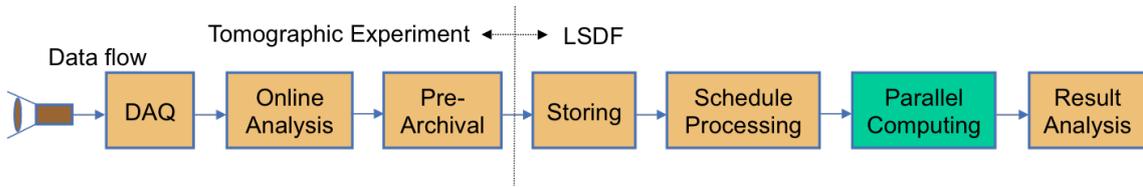
Fig. 1.8. An automated data management and processing workflow designed for tomographic experiments, showing the dataflow from data acquisition to result analysis.

UFO [34]. It will perform on-line assessment of sample dynamics, providing insight into scientific processes which are hardly accessible. However, the storing and processing of the data from high-speed and high-throughput tomographic imaging are challenging tasks and require intelligent intensive data services.

The Large Scale Data Facility LSDF (located at KIT) is a data service facility dedicated for data intensive scientific experiments [57, 58]. It provides not only nearly unlimited storage for large amount of experimental data but also a connected Hadoop cluster used as a computing resource for applications demanding data intensive computing. Based on the LSDF, an automated workflow for the data management and processing of data from tomographic experimental station is designed. The workflow is shown in Fig. 1.8. It is designed for tomographic experiments with dataflow from data acquisition to result analysis, connecting the experimental station to the Large Scale Data Facility (LSDF). At the tomographic experiment side, it focuses on the online assessment relying on the techniques developed by the UFO project to ensure the speed of data acquisition and the quality of tomographic imaging. At the LSDF side, it aims to properly store and archive the large volume of datasets using related meta data, schedule reconstruction process, and perform advanced computing technique by applying the computing source. The final computed results can be accessed by scientific researchers (such as biologist as a user). This connection of tomographic experiment station and LSDF can greatly enhance the storage, processing efficiency, and result analysis for large amount of data acquired from the tomographic experiments.

The workflow above is comprised of works done by multiple research groups. The topic in this thesis, sparse-view reconstruction, is carried on in the block "parallel computing" highlighted by green color in Fig. 1.8. Even thought iterative reconstruction method has been well developed based on the compressive sampling theory for the sparse-view tomographic reconstruction and numerous advanced algorithms have been presented for solving the mathematically modeled problem such as the optimization in Eq. 1.5 and 1.6,, challenges are still existing in the automated workflow for data management and processing. It requires the reconstruction method applied in this workflow to be automated and near real time implemented. This, however, is prevented by the manual operations and intensive computing needed in the reconstruction algorithm.

A critical factor which challenges the automated implementation is the trade-off parameter λ introduced when the constrained minimization problem in Eq. 1.6 is transformed to an unconstrained problem in Eq. 1.5 by using the Lagrangian multiplier method. Even more parameters other than λ can be introduced by more complicated techniques such as augmented Lagrangian technique. These parameters control the balance of data fidelity (minimum error) and regularization (prior knowledge), and are generally difficult to determine.

The determination of these parameters by reasonable, additional conditions poses a multidimensional optimization problem which is computationally expensive. It is interesting to discuss a particular example called TVAL3 proposed in [51]. There regularization is generalized in terms of a variety of constraints formulated by new variables ω_i in addition to $\mathbf{x}$, giving rise to a new minimization problem in ω_i and $\mathbf{x}$. The formal substitution of these additional constraints into the minimization problem gives back the higher dimensional equivalent to the original problem of Eq. 1.5. The important insight is to treat the additional constraints also in the sense of assigning Lagrangian multipliers $\nu_{i,1}$. Consequently, higher powers of constraints are included into the new objective function, subject to additional multipliers $\mu_1, \mu_2, \cdots, \nu_{i,1}, \nu_{i,2}, \cdots$, allowing the approach to the solution of the constraint equations. Now the parameter determination is to find the value of the multiplier set, in a high-dimensional space spanned by all residuals. Practically, this is prohibitive in view of computational effort and high-dimensional space.

The optimal values of these parameters vary for different data sets. A set of badly selected values for these parameters will result in a low quality reconstruction. Pragmatically, one may just set these parameters to values which visually lead to somewhat expected results. However, the manual parameter determination by human judgment should be avoided in the automated workflow of tomographic reconstruction (Fig. 1.8).

Another challenge for tomographic reconstruction comes from the intensive computing required by the algebraic reconstruction technique. Enormous amount of operations on large matrices like multiplication are needed in the forward and backward processes, leading to highly time-consuming reconstructions. The sequentially implemented volume reconstruction will take tens of hours. So parallel computing is preferred to achieve near real-time reconstruction of 3D volumes.

1.5. Proposed methods

The proposed method of this thesis is to reconstruct the tomographic data in a precise and automated way in the case of limited number of projections in order to reduce the X-ray radiation dose or adapt the imaging configuration. By looking into the related work in the

literature, the iterative methods are usually applied for reducing artifacts in the reconstruction of limited projection data by integrating additional information. The forward model is firstly built up for calculating the system matrix and constructing the equation system based on the geometry of the X-ray beam and approximations made for the imaging process. An accurate forward model can well control the data consistence and thus improve the reconstruction image quality. In order to constrain the artifacts caused by the missing angular projections, the sparsity in the gradient transform domain is taken into account, forming the minimization problem in (1.5) by integrating the total variation regularization for iterative reconstruction. A simple implementation of the conjugate gradient method is pursued to solve the unconstrained minimization problem in (1.5) for a yet undetermined parameter λ.

A method for precisely and automatically determining the optimal value of the parameter λ is required for a high quality reconstruction and fitness to the designed workflow in Fig. 1.8. In solving problem (1.5) in the exact sense of problem (1.6) the minimum $\mathbf{x}^*(\lambda)$ would have to be inserted into the minimization condition to yield the optimal value λ^* by algebraic inversion. As a consequence, $\mathbf{x}^*(\lambda^*)$ solves problem (1.6). To proceed like this, however, is impractical since (i) the determination of λ^* requires an iterative approach, e.g., Newton's method, (ii) λ^* is usually not unique implying the need for manual selection, and (iii) a solution in this sense may not exist at all for inconsistent data. Even if it existed, the search for an exact solution to problem (1.6) may be too restrictive in the sense that reconstructions may ensue that are unrealistically flat.

Therefore, a search for a *reasonable* constraint is carried out for fixing the value of λ in (1.5) to a finite value in an automated way. This can be achieved by the celebrated discrete L-curve method [59] which will be employed in the present thesis. More complex parameter selection methods are proposed in the literature such as the discrepancy principle (DP) [60] and the generalized cross validation (GCV) [61]. For an elaborate comparison of the L-curve method with other ways of fixing model parameters in a prior under-determined inverse problem, see [62]. As a result, a parameter-optimized reconstruction algorithm CGTV (total variation-based conjugate gradient) is developed for computed tomography with limited projections, enabling an automated data management and analysis workflow in Fig. 1.8.

The expensive computing required by the algebraic reconstruction of a full 3D volume is greatly alleviated by parallel volume reconstruction through a computer cluster which is directly connected to the Large Scale Data Facility (LSDF). The 3D volume reconstruction is decomposed into numerous slice reconstruction tasks executed in parallel, the outputs of which are later stacked to form the 3D volume. The computing time is reduced from tens of hours to several tens of minutes, achieving near real-time processing.

1.6. Main contributions

The developed method in this thesis allows a high quality reconstruction and a completely automated workflow for the 3D computed tomographic reconstruction of limited number of projections. The precise forward model, regularization based on sparsity assumption, and properly selected value of the parameter λ ensure the high quality reconstruction voiding the serious artifacts generated in the FBP reconstruction for insufficient projections and preserving features in the object as well. Meanwhile, the application of the L-curve based CGTV method successfully avoids the manual operation in parameter adjusting based on an automated selection strategy in the optimization problem of reconstruction. The presented parameter-optimized iterative reconstruction method, L-curve based CGTV, combines the conjugate gradient method, the total variation, and the L-curve method for the reconstruction for limited projection data.

For the first time the developed method and automated workflow are applied to three tomographic data sources: X-ray tomographic imaging at the TOPO-TOMO bending-magnet beamline of ANKA, X-ray tomographic imaging at the undulator beamline 32-ID of APS, and electron tomographic imaging in nanoscopy. In later chapters, it is shown that the parameter optimized algebraic reconstruction method is general for various experimental datasets from different forms of tomography. The method improves the reconstruction quality compared to the traditional reconstruction method and is more flexible than other advanced iterative methods for limited data reconstruction.

The data-parallel computing using the computer cluster at the KIT Large Scale Data Facility (LSDF) allows an accelerated implementation of 3D CT reconstruction. The parallel structure speeds up the whole reconstruction procedure by around 120 times, and finally the near real-time reconstruction is achieved.

As a result, this automated and fast workflow enables the fast tomographic imaging at ANKA, constrains the X-ray radiation dose in developmental biology increasing the duration of *in vivo* time-lapse series, and allows more accurate analysis of nanostructured materials in electron tomography.

1.7. Thesis overview

The thesis is organized in seven chapters as follows:

In the present chapter 1, the requirements for a sparse-view tomographic reconstruction has been motivated. By giving the state of the art it has been seen what is still missing and challenging for our applications. At last, the proposed methods have been briefly described and the main results are presented.

Chapter 2 illustrates the theoretical background of the tomographic imaging method used in this thesis. Concepts, theorems, and models involved in the forward imaging process of computed tomography are covered.

Within chapter 3, a systematic study of tomographic reconstruction methods are carried out. The reconstruction methods are categorized into different groups according to the nature of the algorithms. The comparison indicates the desirable characteristics of the reconstruction algorithms for different situations.

The situation of computed tomography reconstruction with limited number of projections is discussed in chapter 4 based on the compressive sampling theory. The sparsity of the gradient image integrated in the total variation regularization is taken into account for a high quality reconstruction voiding artifacts. A method named CGTV is developed in solving the minimization problem for reconstruction.

Chapter 5 further develops the CGTV reconstruction method by integrating the L-curve method for the determination of the trade-off parameter in order to generate a good reconstruction and build up an automated workflow for tomographic data processing and analysis. The optimized CGTV reconstruction is evaluated with simulated images as well as experimental datasets of living biological objects (weevil and frog embryo).

Chapter 6 addresses the implementation of the 3D tomographic reconstruction which requires intensive computing due to the large amount of tomographic data and the expensive iterative reconstruction method. It describes the data-parallel reconstruction framework and the designed automated workflow for improving the data processing efficiency in various tomographic applications.

Chapter 7 provides a discussion of the covered topics and the results.

Chapter 8 concludes the work in this thesis and gives several recommendations for the future work.

2. Forward problem in computed tomography

The aim of computed tomography (CT) is to obtain information regarding the exact position occupied by the nature of material inside an object. Generally speaking, the process is as follows. A CT scanner is used to produce projections for a specified cross-section of the object by applying X-ray imaging technique. These projections comprise a sinogram, from which a two dimensional image representing the X-ray attenuation index distribution in the cross-section will be produced. The former step, producing projections, is often called a forward imaging process, and the later one is reconstruction process. In practice the projections are always generated by the CT scanner for a whole object at one time instead of only a cross-section.

In this chapter the concepts and theorems involved in the forward imaging process of computed tomography will be described. The intention of this chapter is to give supports to the reconstruction methods described later. In the following fundamentals of tomographic imaging will be introduced, from X-ray imaging to tomographic imaging, from Radon transform, Fourier-slice theorem to forward models and the brief idea for reconstruction. Of course, the theorems for computed tomography can be also found in other popular books for computed tomography. In this chapter the requirements of the reconstruction in this thesis will be also highlighted. The reconstruction of an object from a series of projections enables us to display and discover the precise shape and position of selected targets, which will be discussed in the next chapter.

2.1. X-ray imaging

Prior to introduce any theories of computed tomography, it is necessary to start from introducing the X-ray imaging technique, considering that the physical characteristics of X-ray is the basis for the mathematical expression of the projection obtained by the X-ray. We now describe one-dimensional expression of a projection determined by the X-ray imaging theory.

X-ray imaging is based on the Lambert-Beer law [63, 64] which specifies the dependence of the attenuation index of the X-ray on the properties of the material through which the
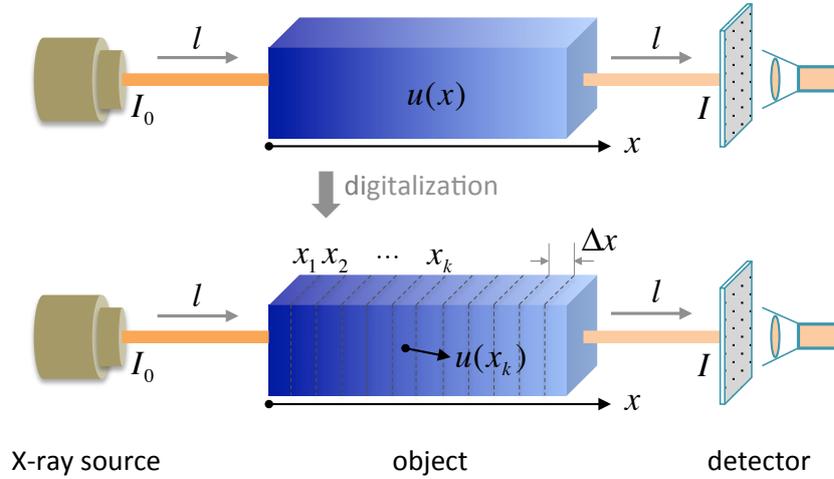
Fig. 2.1. X-ray imaging based on Lambert-Beer law.

X-ray is traveling. As shown in Fig. 2.1, X-ray photons are emitted by the X-ray source and travel through the object to the detector in a modeled straight line l, which is also referred to as the X-ray path. $u(x)$ denotes the X-ray attenuation index distribution of the object along l. In a vacuum all X-ray photons that leave the source in the direction of the detector will reach the detector. There is no attenuation in the traveling process. When an object is placed between the source and the detector, some of photons will be removed from the beam (absorbed or scattered), reducing the possibility of a photon reaching the detector. Thus in this process, the photons are attenuated by the materials in the object and the remains finally reach the detector, leading to a difference between the input and output photon intensities $(I_0,\ I)$. According to the Beer-Lambert law, the input and output photon intensities $(I_0,\ I)$ have the following relationship:

$$I = I_0 e^{-\int_l u(x)\mathrm{d}x}, \tag{2.1}$$

where the line integral of $u(x)$ is performed along the X-ray path l. One can define the projection p of the object as

$$p = \int_l u(x)\mathrm{d}x = \ln \frac{I_0}{I}, \tag{2.2}$$

which is exactly the line integral in (2.1). By measuring the ratio of the intensities I_0 and I using X-ray detector, the projection p can be calculated easily. In Fig. 2.1 only the 1D imaging case is demonstrated. In the 2D or 3D case, the attenuation index as well as the line integral need to be adapted; the attenuation index in 2D case would be $u(x, y)$, in 3D case $u(x, y, z)$.

For building a practical computer expression, the X-ray attenuation index distribution $u(x)$ is generally digitized as a series of finite elements $u(x_k)$, $k = 1, 2, ..., K$, each of which represents the average attenuation index in a fixed interval Δx. Then the line integral for the projection p in (2.2) is reformulated to

$$p = \sum_k u(x_k)\Delta x \,. \tag{2.3}$$

When $\Delta x \to 0$, Eq. 2.3 reduces to the line integral expression in Eq. 2.2. The reconstruction problem now is to estimate a series of discrete values representing the attenuation index $u(x_k)$.

2.2. Concepts for computed tomographic imaging

An interesting problem consists of estimating the attenuation index distribution $u(x)$ from the detected projection data p. It is an inverse process of the forward imaging expressed by Eq. 2.2, commonly referred to as reconstruction. Note that only the measurement in one direction is available for estimating attenuation index $u(x)$ in Eq. 2.2. It is not possible to determine the solution for the inverse procedure. This problem is solved by the so-called computed tomography (CT) technique which applies projections in numerous views.

Now we switch to 2D domain for computed tomography. The principle of tomographic imaging can be easier understood by observing a semitransparent object as shown in Fig. 2.2. Light crosses the semi-transparent object, attenuated by the non-transparent, inner targets. This can be identified by the observer's eyes on the other side. If the observation of the object is performed in the view as displayed in (a), the two targets in the mid-position overlap each other completely so that the true number of targets can not be perceived. If the relative positions of the object and observer are changed as in (b), where the object is rotated by 15 degrees, the overlap is reduced so that the observer can distinguish more targets. If further observations are performed at other angles, the size, density, and number of these targets can be determined precisely.

The illustration in Fig. 2.2 only gives a brief concept of tomographic imaging. The strict mathematical expression for CT forward imaging process was given by the Radon transform, which will be discussed in the next section. This illustration indicates a method for how to recover the inner structure of an object without cutting it. Similarly, a lot of objects are semitransparent for X-rays due to X-rays having strong penetrating ability, such as bones and soft tissues of biological body having different attenuation index for X-rays. A typical data collection method for CT is to posit a X-ray source and a detector on two opposite sides of the object and rotate their relative positions around the center of the object to record data from numerous angles. These recorded data are often called
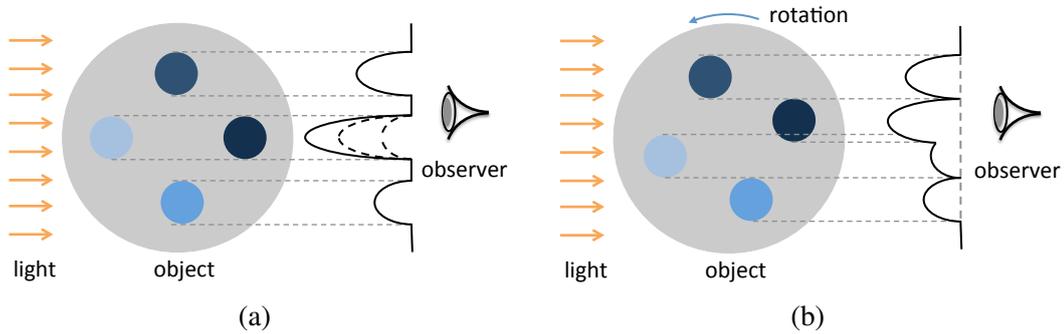
Fig. 2.2. A schematic diagram for understanding the computed tomographic imaging.

measurements. Each measurement is related to an X-ray source position combined with an X-ray detector position. The detector records the number of photons during the exposure time, which is proportional to the X-ray intensity. Thus it is possible to calculate projections from these measurements by using the Lambert-Beer law in Eq. 2.2. Note that the interesting object region in the plane to be imaged should be completely covered by the X-ray beam, otherwise errors can be added in the reconstruction process resulting in bad quality reconstruction image.

The main component of the CT system used for data collection is the CT scanner, comprised of a pair of X-ray source and detector. In the development history of computed tomography, the technology of CT scanners has been improved dramatically and three classical scanning modes had been used by different generations of CT devices. While more variants of these scanning modes existed, we restrict our attention to these three basic modes, which are parallel beam, fan beam, and cone beam, as shown in Fig. 2.3.

The parallel beam is modeled as a linear X-ray source and a linear detector in a plane, where all X-rays go from the source to the detector in parallel. It is the simplest scanning geometry and still utilized in many scientific research fields. Fan beam is modeled as a point X-ray source and a curved detector in a plane, where all X-rays start from one point and radially go to a curved detector. The fan-beam is not of great interest any more in modern CT systems. For the cone beam imaging geometry, the X-rays also start from a point source, but travel in multiple planes to a curved plane detector. Cone beam is a more common used geometry in micro CT systems since it can accelerate the imaging process and it is a real 3D imaging method. However, the simple parallel beam is still the main concern in this thesis, as the experimental data were obtained by parallel X-rays.

The most commonly used representation way for projection data is the so-called sinogram, defined with respect to the projection direction denoted by θ and the radius axis r. For a parallel beam, projections arranged from 0° to 180° contain sufficient information for understanding the object, since additional projections in [180°, 360°] are symmetric repeat of the result in [0°, 180°] considering that X-rays at the angle θ and $\theta+180°$ go
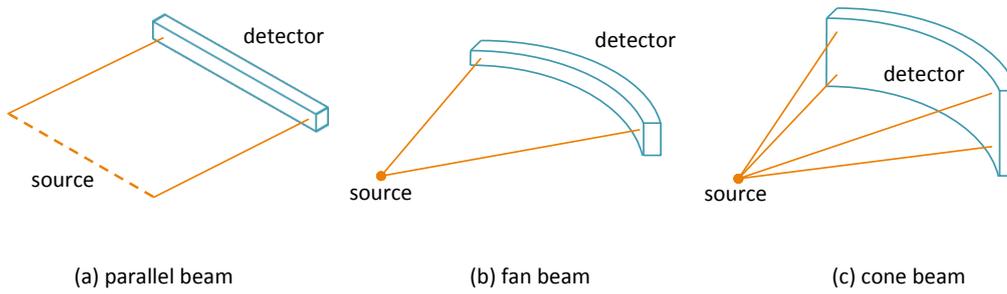
Fig. 2.3. Three types of scanning geometry of computed tomography: (a) parallel beam, (b) fan beam, and (c) cone beam.
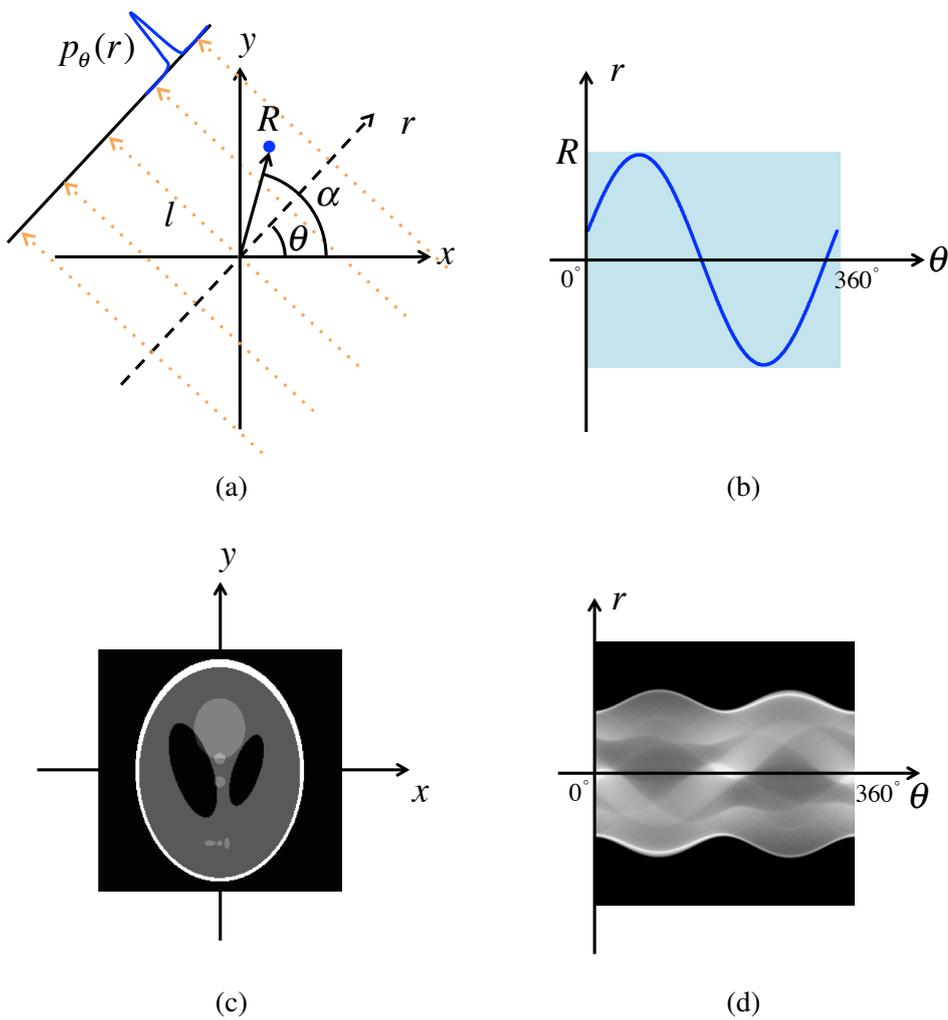
Fig. 2.4. Sinogram of a point with small size in an object. In (a), it shows the projection of a point, denoted by the radial coordinate R, α, leading to a sine-function curve in the sinogram of (b). The sinogram of the Shepp-Logan phantom in (c) is given in (d), which consists of a set of crossed sine-function curves.

through the object along exactly the same path. Different from the parallel beam, the fan and cone beam require tomographic imaging in the full view from 0° to 360°. To show a complete curve, the projection data in [0°, 360°] produced by a parallel beam are considered. In the sinogram, the horizontal axis represents the angle θ in which the measurement is performed. The vertical axis is the distribution of projection $p_\theta(r)$ in a given angle θ. Let us firstly observe the projection of a fixed point in the object, with radial coordinate (R, α) as shown in Fig. 2.4a. For $\theta = [0°, 360°)$, the position of its projection on the axis r satisfies

$$r = R \times \cos(\alpha - \theta) \,. \tag{2.4}$$

So the projections of a point in the sinogram form a sine function as displayed in Fig. 2.4b. As any object can be regarded as a set of points distributed in the spatial domain, we could expect that its sinogram consists of some crossed sine-function curves. This can be seen from the sinogram in Fig. 2.4d, which is obtained from the projections of Shepp-Logan phantom in Fig. 2.4c.

The sinogram is an effective method in detecting unusual issues in the tomographic system. A defective pixel in the detector, for example, results in a straight line in the sinogram, because a single pixel in the detector has the same distance to the center in all angles.

2.3. Radon transform

The strict mathematical expression for CT forward imaging process was given by the Radon transform [2, 65] which can be traced back to 1917. The Radon transform, for the first time, elaborated the imaging theory of computed tomography mathematically, claiming that the original object can be reconstructed from infinite projections in different angles. The Radon transform in 2D continuous Euclidean space is discussed below.

The 2D Radon transform relates a 2D function to the collection of line integrals of that function. As shown in the left panel of Fig. 2.5, $f(x, y)$ denotes the two-dimensional, continuous distribution of the attenuation index of the object in the coordinate system (x, y). The X-rays go through the object along the path l in the direction denoted by the angle θ, generating the projection $p(\theta, r)$ in a radial coordinate system (θ, r). The 2D Radon transform of $f(x, y)$ is defined as

$$p(\theta, r) = \Re f(x, y) = \int_l f(x, y)\mathrm{d}x\mathrm{d}y \,, \tag{2.5}$$

which has the same form as the Lambert-Beer law. Consider the coordinate system (r, s) as a rotated version of the original coordinate system (x, y) as shown in the left panel of
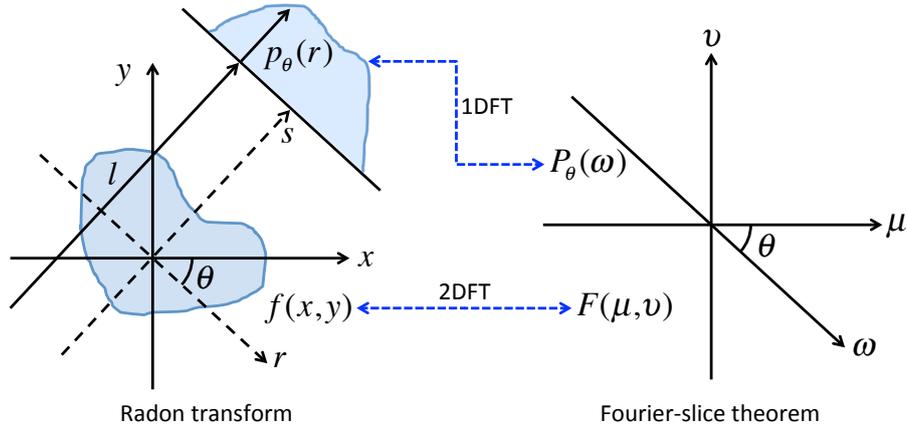
Fig. 2.5. Two-dimensional Radon transform and Fourier-slice theorem. Left panel is the 2D Radon transform of a continuously distributed object $f(x, y)$, generating the projection $p_\theta(r)$. Right panel shows the Fourier-slice theorem giving the relationship of $f(x, y)$ and $p(\theta, r)$ in Fourier domain.

Fig 2.5. Their relationship is expressed as

$$\begin{bmatrix} r \\ s \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}. \tag{2.6}$$

The θ shown in Fig 2.5 is negative. Then the Radon transform can be reformulated to

$$\Re f(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\delta(x\cos\theta + y\sin\theta - r)\mathrm{d}x\mathrm{d}y = p(\theta, r), \tag{2.7}$$

where the function $\delta(\cdot)$ is 1D Dirac function which samples $f(x, y)$ on the straight line $x\cos\theta + y\sin\theta = r$. Note that the 2D Radon transform is the projection of $f(x, y)$ defined as a function of variables θ and r. The inverse process is to look for an operator $\Re^{-1}$ so that for all points (θ, r) we have

$$\Re^{-1}p(\theta, r) = f(x, y). \tag{2.8}$$

The formula for this operation was proved by Radon [2, 4] indicating the method to solve the reconstruction problem. However, in practice some difficulties arise while applying the mathematical solution to the idealized CT problem due to the finite set of measurements and their inaccuracies introduced in the CT imaging process. Algorithms, which can run efficiently in a computer, are needed for producing acceptable reconstructions in spite of the finite and inaccurate nature of the data.

2.4. Fourier-slice theorem

The Radon transform specifies the relationship of the 1D projection $p(\theta, r)$ with the 2D function $f(x, y)$ in spatial domain. Their connection in Fourier domain is provided by the so-called Fourier-slice theorem [30]. Both the Radon transform and Fourier-slice theorem constitute the foundation of analytical reconstruction methods. Let $F(\mu, \upsilon)$ denote the 2D Fourier transform of the object $f(x, y)$ and $P_\theta(\omega)$ the 1D Fourier transform of projection $p(\theta, r)$ with respect to r. For better understanding, $p(\theta, r)$ is now written to $p_\theta(r)$. Then we have

$$F(\mu, \upsilon) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) e^{-i2\pi(\mu x + \upsilon y)} \mathrm{d}x \mathrm{d}y \tag{2.9}$$

and

$$P_\theta(\omega) = \int_{-\infty}^{\infty} p_\theta(r) e^{-i2\pi\omega r} \mathrm{d}r \,. \tag{2.10}$$

The Fourier-slice theorem states that the 1D Fourier transform (FT) $P_\theta(\omega)$ is a straight line through the 2D Fourier transform $F(\mu, \upsilon)$ at the angle θ, which is

$$P_\theta(\omega) = F(\omega \cos\theta, \omega \sin\theta) \,, \tag{2.11}$$

with $\mu = \omega \cos\theta$ and $\upsilon = \omega \sin\theta$ in mathematical notation. This is also graphically shown in the right panel of Fig. 2.5.

The derivation of the Fourier-slice theorem can be carried out in the alternate coordinate system (r, s) expressed in the Eq. 2.6, where $f(x, y) \to f(r, s)$. The projection $p_\theta(r)$ in Eq. 2.7 is then represented as

$$p_\theta(r) = \int_{-\infty}^{\infty} f(r, s) \mathrm{d}s \tag{2.12}$$

and its 1D Fourier transform in Eq. 2.10 becomes

$$P_\theta(\omega) = \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} f(r, s) \mathrm{d}s \right] e^{-i2\pi\omega r} \mathrm{d}r$$
$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(r, s) e^{-i2\pi\omega r} \mathrm{d}r \mathrm{d}s \,. \tag{2.13}$$

This can be transformed back to the coordinate system (x, y) by using the relationship in Eq. 2.6, producing

$$P_\theta(\omega) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) e^{-i2\pi\omega(x \cos\theta + y \sin\theta)} \mathrm{d}x \mathrm{d}y$$
$$= F(\omega \cos\theta, \omega \sin\theta) \,. \tag{2.14}$$

Thus the Fourier-slice theorem in Eq. 2.11 is reproduced.

The Fourier-slice theorem denotes a route for the inverse problem, which attempts to retrieve the object $f(x, y)$ by applying 1D FT and 2D inverse FT respectively. It requires 3 steps. First, the 1D FT is taken for each projection. Secondly, these projection spectra are used to calculate the 2D frequency spectrum of the image. Since the projection spectra are arranged radially, and the image spectrum is arranged rectangularly, an interpolation routine is needed to make the conversion. Thirdly, 2D inverse FT is taken of the image spectrum to obtain the reconstructed image. The interpolation error in Fourier domain leads to larger error in the spatial domain, especially for the high frequency components where the radially arranged samples are more sparser. An alternative method applies interpolation in spatial domain. It uses filtered 1D inverse FT instead of 2D inverse FT, which can be more simply calculated and implemented on a computer. This is the so-called filtered back-projection method (FBP). The FBP reconstruction method will be discussed in Sec. 3.2.

2.5. Digitization and forward models

In the previous sections, only the continuously expressed projections $p(\theta, r)$ were discussed theoretically. However, in practical X-ray imaging procedure, the projections are naturally digitized by the detector with respect to θ and r, namely, $\theta \rightarrow \theta_k$ and $r \rightarrow r_s$ as shown in Fig. 2.6, $k = 1, 2, \cdots, P$ and $s = 1, 2, \cdots, S$. Since the purpose of CT is to reconstruct the real object from its actual X-ray projections, the Radon transform in continuous domain as shown in Fig. 2.5 is unfeasible for the reconstruction from computer simulated projection data.

The object is generally digitized to a grid image in a limited, rectangular region centered at the origin of the coordinate system. This region covers the whole object and is subdivided into a set of small squares. These small squares are referred to as image elements called pixel. An assumption behind this subdivision is that the attenuation index of the material in this small region can be regarded as consistent. Now the expression of the object $f(x, y)$ can be rewritten into $X(m, n)$ or X_{mn} ($m = 1, 2, \cdots, M$ and $n = 1, 2, \cdots, N$), as shown in Fig. 2.6, where M and N are typically but not necessarily equal to each other. Identical object functions may give rise to different images if the image regions are different. Here it is required that the image region should be large enough to cover the original object, with assuming that the value outside the image is zero. In CT, the grid image region is the reconstruction region and the gray value of the image represent the attenuation index of the object at corresponding position.

Note that, from now on, X is used for an image pixel representing the attenuation index, x for x-axis or a variable, $\mathbf{x}$ employed later for a vector representation of all pixels.
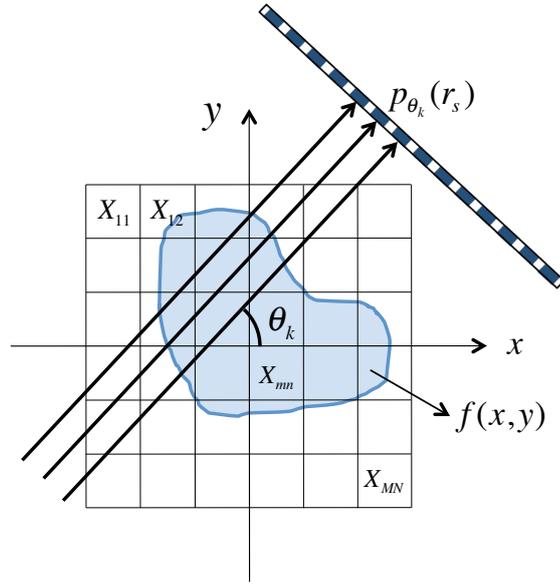
Fig. 2.6. Digitization of the object to a grid image in a limited, rectangular region, covering the whole object to observe.

Now the question is how to connect the computer simulated object $X(m,n)$ to the projections $p_{\theta_k}(r_s)$. This connection is known as forward transform, representing the forward imaging procedure of the CT system. This forward transform can be written into an operation/function, which does not demand too much memory but requires more computing time if the operation is performed many times. Another option is to express the forward transform as a product of a large, sparse matrix called system/transform matrix and the object elements $X(m,n)$, which is commonly used in the iterative reconstruction methods. The transform matrix consists of large amount of weighting factors representing the contribution of each pixel to the measured projections $p_{\theta_k}(r_s)$. For all m and n, we write the object $X(m,n)$ into a vector $\mathbf{x}$ with elements x_i ($i = 1, 2, \cdots, C$, $C = M \times N$ represents the whole number of the image cells). For all k and s, $p_{\theta_k}(r_s)$ is written into a vector $\mathbf{p}$ with elements p_j ($j = 1, 2, \cdots, R$, $R = P \times S$). Let $\mathbf{A}$ denote the transform matrix. The forward problem in CT system is expressed as

$$\mathbf{A}\mathbf{x} = \mathbf{p}. \tag{2.15}$$

The matrix $\mathbf{A}$ can be calculated only once and used for numerous times. The calculation of the matrix $\mathbf{A}$ is generally derived from the so-called forward model or projection operation. The error introduced by the forward model apparently has an important influence on the quality of reconstruction results. The better the forward projection models the imaging process, the better the estimated image matches to the measured data. So a precise forward model is always preferred to pursue a high-quality reconstruction.

There exist already many methods to simulate the physical system or acquisition process in the literature . Most of them can be subdivided into three categories: pixel-driven [66], ray-driven [27, 67] and distance-driven approaches [68, 69]. The pixel-driven projection works by connecting the ray line based on the center of image cells. It is suitable for hardware implementation and more often used in filtered back-projection method. This approach is rarely applied as it has been proved that it introduces high-frequency arti-facts in projections [70, 71]. The iterative reconstruction methods commonly apply the ray-driven approach which is easier to implement for different imaging geometries. It determines the ray line by the detector center instead. The distance-driven approach was proved to be the most accurate one in modeling the forward transform process. However, a complicated forward model usually becomes more time-consuming in calculation. One can select the proper forward model according to the practical requirements in the case of interest.

As the concern of this thesis is mainly focused on the algebraic reconstruction methods we now follow the ray-driven approach and give four choices for the forward model with different assumptions taken for the X-rays. The ray-driven forward model traces rays through the grid image and approximates each line integral as a weighted sum of all image pixels. The rays are generally modeled as a straight, thin line without width as shown in Fig. 2.7a and 2.7b. However, a more precise approximation of the X-rays is the 'fat' lines as shown in Fig. 2.7c and 2.7d which consider the geometry of the detector elements. Especially in Fig. 2.7d the distance between two adjacent rays is also taken into account compared to the 'fat' lines in Fig. 2.7c.

Now let us define the weighting factor a_{ij} representing the contribution of the ith cell x_i to the jth projection p_j. Define the size of each image cell as unit. In Fig 2.7a, a_{ij} is simply defined as one or zero in the way that if the jth ray line goes through the cell x_i, $a_{ij} = 1$, otherwise $a_{ij} = 0$. This is a rough estimation for the weighting factor, even though achieves simple and fast implementation in computer. The second way to calculate a_{ij} is demonstrated in Fig. 2.7b that a_{ij} is equal to the intersection length of the ray and image cell x_i. $0 \leq a_{ij} \leq \sqrt{2}$. This model makes more sense since it apparently distinguishes the contributions of x_i and x_{i+1} to the projection p_j, marked by the blue and orange line respectively. It achieves a good trade-off between accuracy and computation complexity. But more accurate forward models are to take into account the geometry of the detector pixels, so that the rays have certain width as in Fig. 2.7c and also distance as in Fig. 2.7d. The weighting factor a_{ij} is defined as the intersection area of the ray and the image cell x_i, as marked in blue for both cases. $0 \leq a_{ij} \leq 1$. They are more accurate forward models but at the expense of computing time.
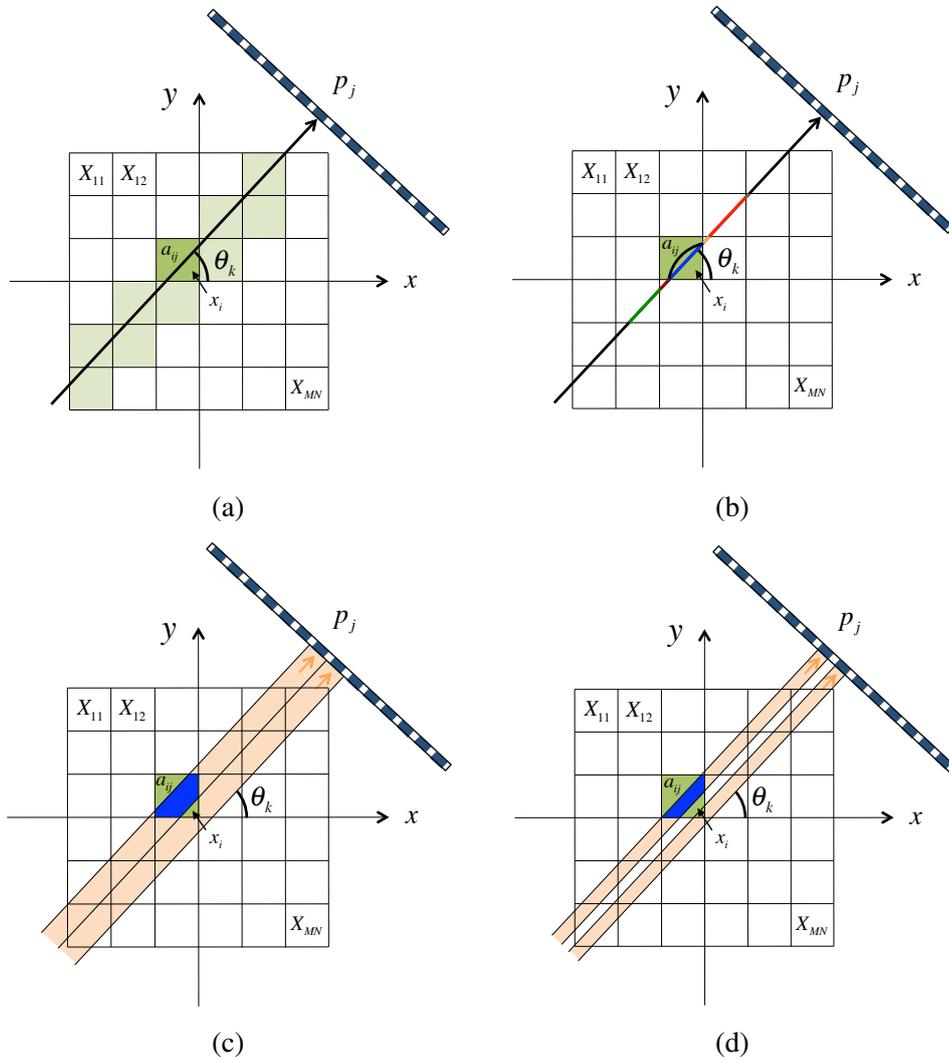
Fig. 2.7. Four forward models based on ray-driven approach for parallel X-ray beam.

Once the weighting factor a_{ij} is calculated for all $i = 1, 2, \cdots, C$ and $j = 1, 2, \cdots, R$, we have the following equations:

$$
\begin{cases}
a_{11}x_1 + a_{21}x_2 + \cdots + a_{C1}x_C = p_1 \, ; \\
a_{12}x_1 + a_{22}x_2 + \cdots + a_{C2}x_C = p_2 \, ; \\
\quad\quad\quad\quad\quad\quad\quad\vdots \\
a_{1R}x_1 + a_{2R}x_2 + \cdots + a_{CR}x_C = p_R \, .
\end{cases}
\tag{2.16}
$$

It can also be written into a matrix form

$$
\begin{bmatrix}
a_{11} & a_{21} & \ldots & a_{C1} \\
a_{12} & a_{22} & \ldots & a_{C2} \\
\vdots & \vdots & \ddots & \vdots \\
a_{1R} & a_{2R} & \ldots & a_{CR}
\end{bmatrix}
\begin{bmatrix}
x_1 \\ x_2 \\ \vdots \\ x_C
\end{bmatrix}
=
\begin{bmatrix}
p_1 \\ p_2 \\ \vdots \\ p_R
\end{bmatrix} .
\tag{2.17}
$$

If we let

$$
\mathbf{A} =
\begin{bmatrix}
a_{11} & a_{21} & \ldots & a_{C1} \\
a_{12} & a_{22} & \ldots & a_{C2} \\
\vdots & \vdots & \ddots & \vdots \\
a_{1R} & a_{2R} & \ldots & a_{CR}
\end{bmatrix} ,
\tag{2.18}
$$

the forward equation system of CT in Eq. 2.15 is achieved. Note that for each projection p_j, the ray only goes through a small number of image cells, which leads to numerous zeros for a_{ij}. Thus the transform matrix $\mathbf{A}$ in Eq. 2.18 is very sparse, greatly reducing its requirement to the memory space.

In this thesis, the forward model in Fig. 2.7b is generally used since the line length is easier and faster to calculate compared to the area in Fig. 2.7c and 2.7d, and meanwhile keeps the accuracy in a high level. The method for calculation of the line length can refer to [27].

2.6. Summary

In this chapter, the forward problems involved in computed tomography, from X-ray imaging to tomographic imaging, from Radon transform, Fourier-slice theorem to sinogram and forward models, are introduced.

The Beer-Lambert law gives the basis for the X-ray imaging, deriving the operation of line integral for X-ray projection. It also applies to the tomographic imaging where Radon transform, for the first time, mathematically elaborated the imaging theory of computed tomography, claiming that the original object can be reconstructed from a set of

projections in different angles. The domain of projection data obtained from the Radon transform is commonly represented by the sinogram which consists of a set of overlapped sine-function curves. The relationship between the original object and its projections performed by Radon transform is revealed by the so-called Fourier-slice theorem in Fourier domain. Both the Radon transform and Fourier-slice theorem constitute the foundation of analytical reconstruction methods.

However, the main concern of this thesis is on the iterative reconstructions, where the forward models are playing an important role. The forward model connects the computer simulated object to the projections. This can be considered to be the discretization process of the Randon transform and represents the forward imaging procedure of the CT system. For the iterative reconstruction method an equation system is built up through the system matrix. The elements of the system matrix denotes the contribution of each pixel to the projections and can be calculated from different forward models. A precise forward model is supposed to improve the quality of reconstruction result by avoiding introducing errors. But an accurate, complicated forward model usually becomes more time-consuming in calculation of the system matrix. Thus one should select a proper forward model according to the practical requirements in the case of interest.

3. Computed tomographic reconstruction

With this chapter a systematic study of reconstruction algorithms is made. The purpose of computed tomography (CT) reconstruction is to restore the 2D/3D distribution of the attenuation index of the object from the measured, tomographic projections. There already exist numerous of reconstruction methods in the literature, which can be generally divided into two groups: analytical reconstruction methods and iterative reconstruction methods. The nature of the algorithms will be explained in the two groups in order to indicate the desirable characteristics of different reconstruction algorithms.

3.1. Problem statement

Until now $f(x, y)$ in Cartesian coordinate system has been used to denote the attenuation index distribution of an object at the point (x, y). By digitization in Sec. 2.5, a square, grid image X_{mn} has been defined to represent the object. The image region encloses the object and the value outside the image region is zero. For convenience, we rewrite the image X_{mn} into a vector formulation $\mathbf{x}$. For projection data, three corresponding expressions have been applied, that are $p(\theta, r)$ for a continuous expression, $p(\theta_k, r_s)$ or $p_{\theta_k}(r_s)$ for actually measured projections, and $\mathbf{p}$ for a stacking projection vector. The later two expressions can be employed for the reconstruction in a computer.

When designing a reconstruction algorithm the data collection method is assumed to be fixed and known. Thus the forward model can be designed according to the data collection method such as the scanning geometry. The method for creating the forward model has been illustrated in Sec. 2.5. Roughly stated, the reconstruction problem is to estimate the image $\mathbf{x}$ given the forward model and the produced projection data $\mathbf{p}$. The reconstruction quality can highly depend on the accuracy of the forward model and the projection data. If the forward model does not well reflect the physical imaging process of X-ray CT, or the acquired projection data are contaminated by large amount of measurement errors, we can not expect a good reconstruction correctly displaying inner structures of the object. Another factor that can determine the reconstruction quality is the reconstruction method

employed. Two categories of reconstruction methods are the analytical reconstruction method and the iterative reconstruction method.

Analytical reconstruction methods are based on the analysis in the Fourier domain. The Radon transform and Fourier-slice theorem denote a route for the reconstruction problem, which attempts to retrieve the object by applying 1D Fourier transform (FT) and 2D inverse Fourier transform (IFT) respectively. However, this route theoretically requires infinite number of projections. In practice only a finite number of projections of an object are available. In this case this reconstruction route can result in degraded image due to the interpolation error from the radial coordinates to rectangular coordinates. An alternative method, using filtered 1D IFT instead of the 2D IFT, can be more simply implemented by smearing back the projections to the object image along the imaging straight line. This leads to the so-called filtered back-projection method (FBP). FBP method reconstructs the object only once without iterations needed. As it operates the projections one by one, the FBP reconstruction can be carried out simultaneously with the data acquisition process and performed in parallel. This makes it to be a fast and real time reconstruction process. Thus the FBP method was commonly used in most CT systems since the CT technology was discovered.

Different from the FBP method, iterative reconstruction (IR) methods assume that the object consists of an array of unknowns, for which a linear equation system is build up according to a forward model. IR methods aims to numerically solve this equation system for the unknowns from the measured projections and the forward model. They rebuild the original object using numerous iterations, in which the current estimation repetitively converges towards to a better solution according to certain criterion. The concept of IR was successfully used in the early age of CT [72], but not widely applied due to its higher demands for computation compared to the analytical reconstruction methods. This situation has been changed by the large computational capability available in the normal workstations at present. Various parallel computing techniques in modern processors (central processing unit, CPU) or graphics adapter (graphics processing unit, GPU) are employed for acceleration. The popularity of the modern IR methods has been also attributed to the considerable effort invested to the reconstruction with limited data. The usage of IR is becoming a more realistic, popular option in the present study of tomographic reconstruction.

Figure 3.1 shows the main categories of tomographic reconstruction method. The tomographic reconstruction methods are first classified to the two groups, analytical and iterative reconstruction methods. Classical reconstruction algorithms FBP and FDK [73] belong to this group. Depending on whether regularization method is applied or not, iterative reconstruction methods can be further divided into to different categories. The ART-like and EM-like reconstruction methods in the literature do not include any regularization for the unknown image. They generally obtain the similar reconstruction quality with the FBP method. Both of them are not able to handle the situations when a limited
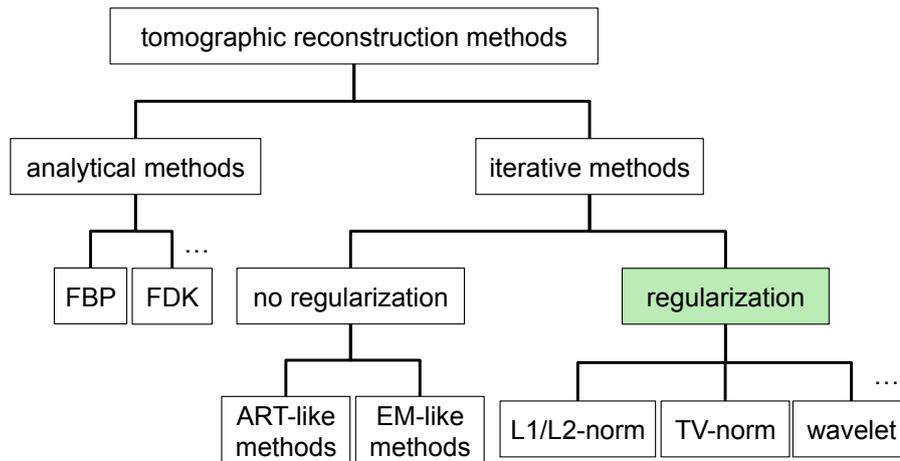
Fig. 3.1. Categories of tomographic reconstruction methods.

number of projections are available or when the obtained projections are not uniformly distributed over 180°or 360°. Problems of this type can be addressed by the iterative reconstruction method with regularization, which incorporate prior knowledge such as sparsity in the spirit of the compressive sampling theory. This chapter will generally discuss these reconstruction methods and highlight the problem of concern, reconstruction with limited projections.

3.2. Analytical reconstruction methods

The primary focus of this thesis is on the iterative methods for tomographic image reconstruction with limited projection data. Nevertheless, analytical reconstruction methods are important when limited computation time is highly required and an approximate reconstruction is tolerable. The fast analytical reconstruction is also often used for the initial guess in the iterative methods to achieve faster convergences. The Radon transform [2, 65] and Fourier-slice theorem [30] build up the foundation for analytical reconstruction methods. As the FBP reconstruction is the most representative and widely used method among the analytical methods, its basic principle and implementation is discussed in the following.

3.2.1. Reconstruction method using inverse Fourier transform

Recall the Fourier-slice theorem expressed by Eq. 2.11 and the corresponding diagram in Chapter 2, repeated here in Fig. 3.2 and the following formula:
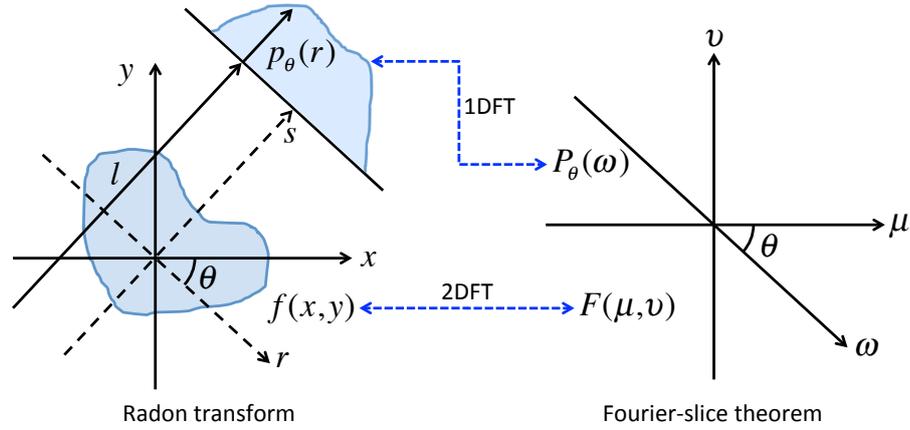
Fig. 3.2. Two-dimensional expressions and configurations for the object $f(x, y)$ and its tomographic projections $p_\theta(r)$ in the spatial domain (left panel) and the Fourier domain (right panel).

$$P_\theta(\omega) = F(\omega \cos \theta, \omega \sin \theta) \,, \tag{3.1}$$

The 2D Fourier transform of the object $f(x, y)$ is denoted by $F(\mu, \upsilon)$ expressed in Eq. 2.9. The Fourier-slice theorem indicates that the values of $F(\mu, \upsilon)$ can be determined by taking the projections at some angles and transforming them into Fourier domain on radial lines as shown in Fig. 3.2. If an infinite number of projections are taken, any point in the $\mu\upsilon$-plane can be exactly determined in this way, thus $F(\mu, \upsilon)$ is known. Then the object function $f(x, y)$ can be recovered by performing the 2D inverse Fourier transform which can be written as

$$f(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F(\mu, \upsilon) e^{i2\pi(\mu x + \upsilon y)} d\mu d\upsilon \,. \tag{3.2}$$

However, only a finite number of projections of the object are practically available. Since the object function $f(x, y)$ can be always bounded by such a square with length L_e that $-L_e/2 < x < L_e/2$ and $-L_e/2 < y < L_e/2$, the 2D IFT in Eq. 3.2 can be written in a summation form as

$$f(x, y) = \frac{1}{L_e^2} \sum_m \sum_n F(\frac{m}{L_e}, \frac{n}{L_e}) e^{i2\pi(mx + ny)/L_e} \tag{3.3}$$

for integers m and n. In practice the energy contained in the Fourier transform components above certain frequency is negligible. It can be supposed that a finite number (K) of Fourier components are known. Thus we can write
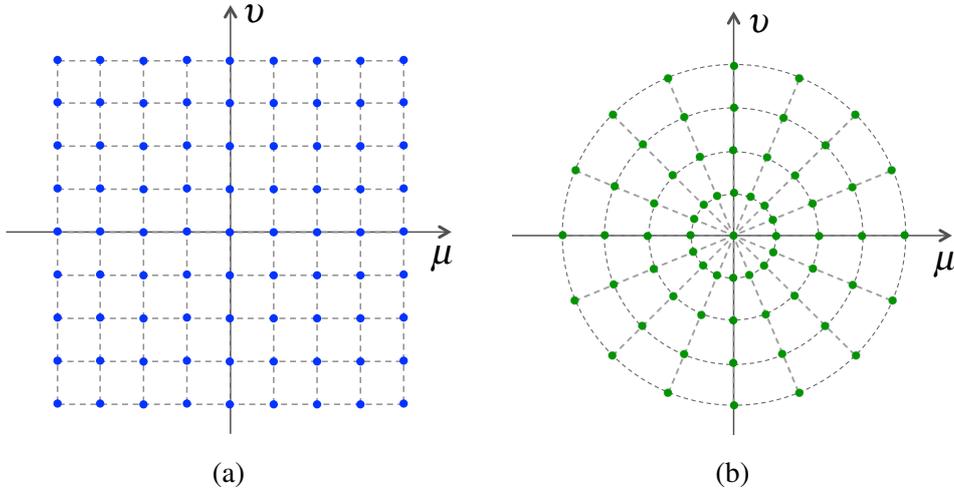
Fig. 3.3. Finite Fourier coefficients at square grid points in (a) and at points along radial lines in (b) determined by Fourier-slice theorem.

$$f(x,y) \approx \frac{1}{L_e^2} \sum_{m=-K/2}^{K/2-1} \sum_{n=-K/2}^{K/2-1} F(\frac{m}{L_e}, \frac{n}{L_e}) \mathrm{e}^{i2\pi(mx+ny)/L_e} , \tag{3.4}$$

where K is arbitrarily assumed to be an even. A larger N will include Fourier components of higher frequencies, which can improve spatial resolution. Equation (3.4) tells that $f(x,y)$ can be approximately determined by the finite K^2 Fourier coefficients $F(m/L_e, n/L_e)$ located at the square grid points as shown in Fig. 3.3a.

However, in practical case the value of $F(\mu, \upsilon)$ is only known along finite radial lines shown in Fig. 3.3b according to the Fourier-slice theorem. In order to perform the inverse Fourier transform in Eq. 3.4, the values of $F(\mu, \upsilon)$ on the square grid must be estimated by interpolation method from the values on these radial lines. The reconstruction result will be degraded due to the interpolation error which becomes greater for higher frequency components since the radial lines are sparser distributed there.

Therefore, even though this idea provides a simple conceptual strategy for tomographic reconstruction, it can not be practically implemented. Instead the FBP algorithm is developed for an accurate and fast implementation by rewriting the inverse Fourier transform in Eq. 3.2 in a polar coordinate and reorganizing the interval of integration according to the Fourier-slice theorem. In the following the derivation of the filtered back-projection algorithm in the context of straight parallel beam for X-ray imaging and its computer implementation are given.

3.2.2. FBP reconstruction method

In the Fourier domain, the Cartesian coordinate system (μ, υ) can be exchanged to the polar coordinate system (ω, θ) by using the relationships

$$\mu = \omega \cos \theta$$
$$\upsilon = \omega \sin \theta$$

$$(3.5)$$

and the differentials

$$\mathrm{d}\mu \mathrm{d}\upsilon = \omega \mathrm{d}\omega \mathrm{d}\theta \,. \tag{3.6}$$

By substituting these expressions in the inverse Fourier transform in Eq. 3.2, we have the new formula in polar coordinate system

$$f(x, y) = \int_0^{2\pi} \int_0^{\infty} F(\omega, \theta) \mathrm{e}^{i2\pi\omega(x \cos \theta + y \sin \theta)} \omega \mathrm{d}\omega \mathrm{d}\theta \,. \tag{3.7}$$

As specified in Sec. 2.2, in the sinogram generated from parallel beam tomography, projections arranged from 0° to 180° contains sufficient information for understanding the object. Additional projections in $[180°, 360°]$ simply repeat the result in $[0°, 180°]$. The reason is that the X-ray beam at angle $\theta + \pi$ passes through the object just in the opposite direction of angle θ, but involves the same line integral path and attenuation coefficients. This makes the projections in these two direction symmetrically equal. Thus in Fourier domain the following property exists

$$F(\omega, \theta + \pi) = F(-\omega, \theta) \,. \tag{3.8}$$

Equation 3.7 can be divided into two parts with respect to θ that

$$
\begin{aligned}
f(x, y) &= \int_0^{\pi} \int_0^{\infty} F(\omega, \theta) \mathrm{e}^{i2\pi\omega(x \cos \theta + y \sin \theta)} \omega \mathrm{d}\omega \mathrm{d}\theta \\
&+ \int_0^{\pi} \int_0^{\infty} F(\omega, \theta + \pi) \mathrm{e}^{i2\pi\omega[x \cos(\theta + \pi) + y \sin(\theta + \pi)]} \omega \mathrm{d}\omega \mathrm{d}\theta \\
&= \int_0^{\pi} \left[\int_{-\infty}^{\infty} F(\omega, \theta) |\omega| \mathrm{e}^{i2\pi\omega r} \mathrm{d}\omega \right] \mathrm{d}\theta
\end{aligned}
\tag{3.9}
$$

with $r = x \cos \theta + y \sin \theta$. According to the Fourier-slice theorem in Sec. 2.4, the Fourier transform of the projection at the angle θ, $P_\theta(\omega)$, is equal to $F(w, \theta)$ so that

$$f(x, y) = \int_0^{\pi} \left[\int_{-\infty}^{\infty} P_\theta(\omega) |\omega| \mathrm{e}^{i2\pi\omega r} \mathrm{d}\omega \right] \mathrm{d}\theta \,. \tag{3.10}$$
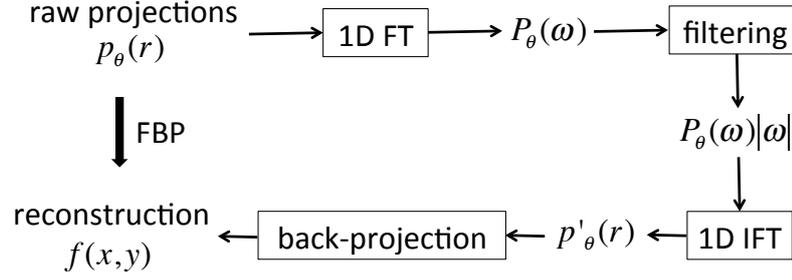
Fig. 3.4. FBP reconstruction can be performed in four steps: (1) one-dimensional Fourier transform, (2) filtering, (3) one-dimensional inverse Fourier transform, and (4) back-projection.

Now we define a new expression $p'_\theta(r)$ as

$$p'_\theta(r) = \int_{-\infty}^{\infty} P_\theta(\omega)|\omega|e^{i2\pi\omega r}\mathrm{d}\omega \,, \tag{3.11}$$

which is the inverse Fourier transform of the term $P_\theta(\omega)|\omega|$, then we have

$$f(x, y) = \int_0^\pi p'_\theta(r)\mathrm{d}\theta \,. \tag{3.12}$$

This estimate of the object $f(x, y)$ is simply expressed as the integral of $p'_\theta(r)$ over different angles θ. $p'_\theta(r)$ is different from the original projection $p_\theta(r)$ in the way that the projection is filtered by $|\omega|$ in the Fourier domain, thus $p'_\theta(r)$ is called a filtered projection. The integral operation in Eq. 3.12 is known as "back-projection". Therefore the whole reconstruction algorithm is called "filtered back-projection".

From Eq. 3.11 and 3.12, it can be seen that the FBP reconstruction is performed in four steps as shown in Fig 3.4. In the first step, the raw projection data $p_\theta(r)$ are transformed into $P_\theta(\omega)$ by one-dimensional Fourier transform. The $P_\theta(\omega)$ is then filtered in the following step, resulting to $P_\theta(\omega)|\omega|$. Thirdly, the one-dimensional inverse Fourier transform is applied to produce the filtered projections $p'_\theta(r)$. Finally the reconstruction image $f(x, y)$ is achieved after carrying out the back-projection operation.

Until now the FBP reconstruction algorithm for continuously expressed projections has been discussed. In practice the recorded projections by the camera are always sampled in discrete form. The discrete expressions for FBP are given below by adding two assumptions and an approximations.

For practical purpose, we assume that the projections are band-limited and the sampling frequency by the camera is high enough for all projections. Namely, we consider that the energy contained in the Fourier transform components above W is neglected. W is

the maximum frequency component for the projections. By the sampling theorem the sampling frequency is $2W$ and the projections are sampled at intervals of

$$T = \frac{1}{2W} .$$
(3.13)

It is also assumed that the projections are space-limited. Namely, the values of projection for large $|r|$ are equal to zero. Thus a projection $p_\theta(r)$ can be expressed as a finite and discrete version

$$p_\theta(mT), \qquad m = -\frac{K}{2}, \cdots, \frac{K}{2} - 1$$
(3.14)

for K sampled values (arbitrarily assume K is an even). Now the Fourier transform of a projection can be approximately written as

$$P_\theta(\omega) \approx P_\theta(m\frac{2W}{K}) = \frac{1}{2W} \sum_{k=-K/2}^{K/2-1} p_\theta(\frac{k}{2W}) e^{-i2\pi mk/K} .$$
(3.15)

Now given the samples of the projections, the samples of their Fourier transform can be obtained by Eq. 3.15. This can be carried out by using one-dimensional DFT. The next step is to obtain the filtered projection $p'_\theta(r)$ digitally instead of Eq. 3.11. Since the Fourier transform $P_\theta(\omega)$ has been assumed to be bandlimited, (3.11) can be approximated by

$$p'_\theta(r) \approx \int_{-W}^{W} P_\theta(\omega) |\omega| e^{i2\pi\omega r} d\omega$$

$$= \frac{2W}{K} \sum_{m=-K/2}^{K/2-1} P_\theta(m\frac{2W}{K}) |m\frac{2W}{K}| e^{i2\pi m(2W/K)r} .$$
(3.16)

Then the sampled filtered projection at $r = k/2W$ ($k = -K/2, \cdots, K/2 - 1$) is given by

$$p'_\theta(\frac{k}{2W}) = \frac{2W}{K} \sum_{m=-K/2}^{K/2-1} P_\theta(m\frac{2W}{K}) |m\frac{2W}{K}| e^{i2\pi mk/K} .$$
(3.17)

By the above equation the samples of the filtered projections are given by the one-dimensional IDFT of the product of $P_\theta(m(2W/K))$ and $|m(2W/K)|$.

The reconstruction image $f(x, y)$ can then be obtained by the back-projection operation in (3.12). One may also approximate this operation as a discrete version to the integral, i.e.,

$$f(x, y) \approx \frac{\pi}{P} \sum_{i=1}^{P} p'_{\theta_i}(\frac{k}{2W}) \,, \tag{3.18}$$

where P is the number of angles at which the projections are sampled. Note that a value of (x, y) may not perfectly correspond to one of the samples of $p'_{\theta_i}(k/2W)$. Thus the discrete back-projection operation requires an approximation by a suitable interpolation algorithm. Often linear interpolation is adequate.

Therefore, the four steps of FBP reconstruction in Fig. 3.4 have been converted to discrete versions by using DFT, IDFT, and back-projection in (3.15), (3.17), and (3.18) respectively.

3.2.3. Sampling requirements

The approximation in (3.18) assumes that the projections in other angles apart from θ_i $(i = 1, \cdots, P)$ are zero. Thus the accuracy highly depends on the number of sampled projections P. If P is too small, serious streak-line artifacts appear in the reconstruction. The FBP reconstruction method has certain limit on the number of sampled projections. In the following the sampling requirements for FBP is discussed.

Suppose the reconstruction image is included in a unit square area and uniformly sampled into N pixels in horizontal and vertical directions. The spatial sampling interval is written as $T_x = 1/N$. Considering the imaging geometry using parallel beam, the projections are equiangularily obtained in P angles, each of which is sampled into K pixels by the detector. The sampling interval of the detector for each projection is written as $T_p = \sqrt{2}/K$ considering that the projections should cover the whole area of reconstruction image. The angular interval is denoted by $\delta = \pi/P$. The Nyquist-Shannon theory has some requirements on the sampling rates of reconstruction image, projections, and angles, namely, the relationships between N, K, and P.

According to the Nyquist-Shannon theorem the maximum frequency of the reconstruction image is $W = 1/(2T_p)$. As shown in Fig. 3.5 the interval in radial direction in frequency domain is $\omega_p = 2W/K$ and the maximum interval from A to B in angular direction can be apparently written as $\omega_\delta = W \cdot \delta$. As the points become sparser when it goes farther away from the center in the radial direction according to the Fourier-slice theorem, it is required
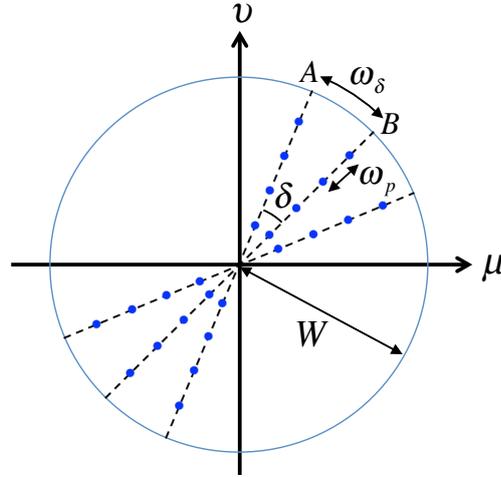
Fig. 3.5. Sampling in Fourier domain

that the resolution in the radial direction is approximately equal to the worst resolution in the angular direction, namely $\omega_p \approx \omega_\delta$. Thus we have

$$P \approx \frac{\pi}{2}K \qquad \text{or} \qquad P > \frac{\pi}{2}K\,. \tag{3.19}$$

Therefore, in FBP reconstruction the number of projections should approximately satisfy Eq. 3.19. Otherwise, the reconstruction image will include streakline artifacts. Besides, in order to keep the resolution existing in the projections, it is necessary to make the the sampling interval of the reconstruction image and that of the projections roughly the same, namely $T_x \approx T_p$. Thus we have

$$K \approx \sqrt{2}N\,, \tag{3.20}$$

which is a common set for the FBP reconstruction.

3.3. Iterative reconstruction methods

The imaging process can be modeled into an equation system in Eq. 2.15. The iterative reconstruction is the inverse problem based on this equation system, which is to reconstruct the object from the measurements. A large number of iterative reconstruction methods can be found in the literature. They are distinguished from each other by the models, iterative strategies, prior information, or solvers applied. Regarding the time when these methods were presented, the iterative reconstruction experiences two important ages.

The early iterative methods were presented for general projections, such as algebraic reconstruction technique (ART) [32], simultaneous algebraic reconstruction technique (SART) [74], simultaneous iterative reconstruction technique (SIRT) [75, 76], maximum likelihood (ML) reconstruction [77, 78], expectation-maximization (EM) reconstruction [79], and ordered subset (OS) reconstruction [80, 81]. Nevertheless, these early iterative reconstruction methods were not widely used in practical applications due to the expensive computation problem. They can only be used for the cases where relatively small amounts of projection data are produced per scan and low-resolution object is reconstructed in the CT system. The common point of these methods is that no regularization was included yet.

In recent years, a large number of studies have been done to improve the quality of reconstruction from limited projection data due to the high demand for the low dose radiation of X-ray and the geometrically fixed configuration of the CT system. The iterative reconstruction methods are suitable for this task as it is easier to cooperate additional conditions. These additional conditions are required since the under-determined/ill-posed system may have multiple solutions in the high-dimensional space or no solutions at all due to conflicts in the measured data. Regularization is a method introducing additional information in order to solve the ill-posed problems. It helps to construct a unique solution to the iterative reconstruction subject to such additional conditions. The reconstruction methods with the regularization are theoretically supported by the compressive sampling theory.

In the following the iterative reconstruction methods which are with and without regularization will be discussed separately.

3.3.1. IR methods without regularization

The iterative algorithms without regularization can be categorized according to their criteria and the ways for updating current images using measured data. Two criteria are generally used: least square and maximum likelihood criterion.

The least square criterion is to minimize the L2-norm function of the data error $\mathbf{Ax} - \mathbf{p}$, which is written as

$$\min \|\mathbf{Ax} - \mathbf{p}\|_2^2, \tag{3.21}$$

It is a typical linear problem. The $\mathbf{x}$ is often solved by using numerical updating method which is called algebraic method.

The maximum likelihood estimation [82, 78] is to maximize the likelihood function which is defined as the possibility of the observations $\mathbf{p}$ given the distribution of object $\mathbf{x}$:

$$L(\mathbf{x}) = P(\mathbf{p}|\mathbf{x}). \tag{3.22}$$
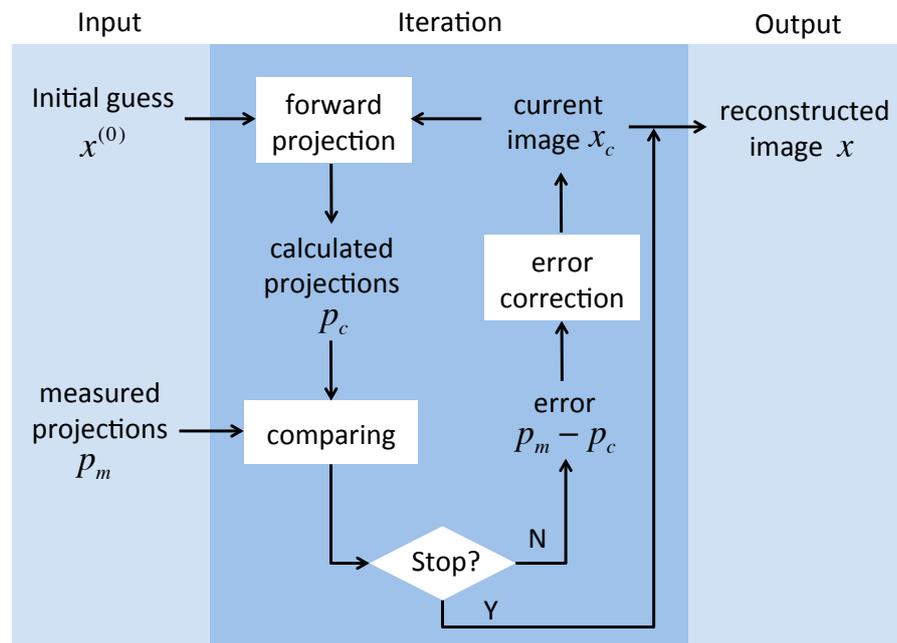
Fig. 3.6. Three major steps of iterative algorithms without regularization: forward projection, comparing, and error correction in the white blocks.

It can estimate the unknown $\mathbf{x}$ given the measured data in the Gaussian or Poisson environments. The ML reconstruction is usually performed by the use of the expectation maximization (EM) algorithm [79], each iteration of which includes two steps: the expectation step (E-step) and maximization step (M-step).

The algebraic reconstruction technique (ART) and expectation maximization (EM) algorithm are the early algorithms widely used due to their simplicity. Later some improve algorithms were designed based on the ART and EM. We call them here the ART-like and EM-like algorithms. In practical implementation, they are proved to have similar updating steps. It can be concluded that these iterative reconstruction algorithms consist of three major steps: forward projection, comparing, and error correction, as demonstrated by the three white blocks in Fig. 3.6. Before starting the iteration, an initial guess of $\mathbf{x}^0$ is made. In the first step, the forward projection is applied to produce artificial projection data $\mathbf{p}_c$ based on the starting point. In the following step, the artificial projection data are compared with the real measured data $\mathbf{p}_m$ to create an error term $\mathbf{p}_m - \mathbf{p}_c$. The error term is then used to correct the estimated object. These three steps are repeated until the stop condition is fulfilled to output the reconstructed image.

ART

The simplest form of the iterative reconstruction is algebraic reconstruction technique (ART) [32]. It was firstly proposed by Gordon etc. [32] and Hounsfield [31] simultaneously by applying the simple procedure proposed by Kaczmarz [83, 84].

The ART algorithm updates the object vector $\mathbf{x}$ using the following error correction:

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \mathbf{A}_j^{\mathrm{T}} \frac{p_j - \mathbf{A}_j \mathbf{x}^{(k)}}{\|\mathbf{A}_j\|_2^2} , \tag{3.23}$$

where $\mathbf{A}_j$ is the combination of all elements a_{ij} in the jth row of $\mathbf{A}$ with $i = 1, 2, \cdots, N$, recalling the expression in Sec. 2.5; $\|\mathbf{A}_j\|_2^2 = \mathbf{A}_j \mathbf{A}_j^{\mathrm{T}} = \sum_{i=1}^{N} a_{ij}^2$.

Note that for one operation of Eq. 3.23 only one measured value of projection p_j is used for updating the estimated object. A complete correction for all measured projections requires M operations of (3.23) constituting one iteration of the whole reconstruction procedure. Thus the ART algorithm is sequential in nature, implementing a correction in such a way that it corrects the error computed from one ray-based equation, then proceeds to another. However, it is likely to exhibit noise in the reconstruction, since data inconsistencies exist in the equation system. The solution is always corrected by a single equation finally, likely leading to a noticeable streak in the particular angle corresponding to that equation. This situation can be alleviated, to some extent, by introducing a relaxation parameter [78] $\lambda^{(k)}$ in (3.23), but obtaining lower convergence property. The updating process with the relaxation parameter $\lambda^{(k)}$ for ART becomes:

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \lambda^{(k)} \mathbf{A}_j^{\mathrm{T}} \frac{p_j - \mathbf{A}_j \mathbf{x}^{(k)}}{\|\mathbf{A}_j\|_2^2} , \tag{3.24}$$

SIRT

Another algorithm which achieves a smoothed reconstruction image, is called the simultaneous iterative reconstruction technique (SIRT) [75, 76]. It improves the ART algorithm by changing the sequential implementation. SIRT applies the same error correction as ART in Eq. 3.23, but keeps the index k invariant in one iteration, so that a series of $\mathbf{x}^{(k+1)}(j)$ with respect to the measured projection p_j $(j = 1, 2, \cdots, M)$ are updated from the same $\mathbf{x}^{(k)}$. At the end of the iteration the new estimated image is calculated by the average of all these updates, which can be written as

$$\mathbf{x}^{(k+1)} = \sum_{j=1}^{M} \mathbf{x}^{(k+1)}(j) . \tag{3.25}$$

This constitutes one iteration of the SIRT algorithm. In the second iteration it goes back to the first equation of the equation system and repeats this process. The average operation in SIRT sets the new estimated image in a proper position which has the same distance to all equations, thus able to result in a smoothed reconstruction, which reduces the noises exhibited in the ART reconstruction.

SART

The algorithm, simultaneous algebraic reconstruction technique (SART), seems to combine the properties of ART and SIRT. It was firstly proposed in [74]. SART updates the object vector by using a variant error correction

$$\mathbf{x}_i^{(k+1)} = \mathbf{x}_i^{(k)} + \frac{\sum_{j=1}^{M} \left[a_{ij} \frac{p_j - \mathbf{A}_j \mathbf{x}^{(k)}}{\sum_{i=1}^{N} a_{ij}} \right]}{\sum_{j=1}^{M} a_{ij}} . \tag{3.26}$$

Note that the SART updates the image cells by simultaneously applying complete measured projections p_j. The total error term is obtained by the weighted average of all error terms with respect to the index j. This average correction to each image cell is calculated and added to generate the updated image. The SART algorithm, on the one hand, maintains the rapid convergence property of ART and offers a smoothed reconstruction as produced by SIRT on the other hand. Employing a relaxation factor can help to further reduce the amplitude of noises.

EM

Apart from the ART-like algorithms, the statistical reconstruction methods were also drawing lots of attention. The expectation maximization (EM) algorithm was firstly presented in a generic form in [79], and has been later used in various tomographic applications [85, 86, 87, 88, 89]. The EM method estimates the original object by computing the maximum likelihood (ML) from the measured data, which is to find the estimated coefficients making the measurements the most probable. It is able to incorporate the statistical properties of the photons in acquisition to the reconstruction process, thus offering reduced noise in the restored object compared to the nonstatistical methods.

OS

The idea of ordered subsets (OS) was proposed to achieve a rapid reconstruction procedure [80, 81]. In OS-based methods, the measured projections are divided into numerous groups called subsets. The error correction is performed according to each subset instead

of the complete projections. By this idea, the convergence speed is improved with respect to the number of subsets.

3.3.2. IR methods with regularization

The prior information in the object can be used to correct the uncertainties in the reconstruction of some practical tomographic data, such as the few-view or limited-angle data. The prior information is often integrated to a regularization in the iterative reconstruction process. The iterative methods with regularization is currently a hot topic in the CT reconstruction [90, 91, 20], aiming to reduce the X-ray radiation dose or adapt the experimental setup.

The commonly used prior information include positivity constraint on the reconstructed attenuation coefficients, piece-wise consistence in spatial domain, and sparse representation in a transformed domain. For example, the author in [20] used the positivity constraint in the iterative process to guarantee the positive property of the attenuation index to estimate. It mapped all negative values to zero after each step of data correction. It also assumed the piece-wise consistence in the spatial domain by minimizing the total variation of the reconstruction image. [92] is an example of applying the sparseness in the wavelet domain of the reconstruction image to gain further improvement in the image quality.

The regularization based on these prior information directs the iteration towards a solution which better fits the measured data and the properties indicated by the prior information. This also leads to a faster convergence. Consequently, the regularization algorithms are able to reduce the noise and artifacts in the reconstruction while preserving the main features and spatial resolution.

The iterative reconstruction method with regularization can be generically written as the following optimization problem:

$$\min_{\mathbf{x}}\{\|\mathbf{A}\mathbf{x} - \mathbf{p}\|_2^2 + \lambda R(\mathbf{x})\}. \qquad (3.27)$$

The first term $\|\mathbf{A}\mathbf{x} - \mathbf{p}\|_2^2$ is again the least square criterion as in (3.21). Here we call it the data fidelity term which makes the reconstructed image $\mathbf{x}$ consistent with the measured data; $R(\mathbf{x})$ denotes the regularization term to incorporate the prior information; the parameter λ controls the trade-off between these two terms. Note that $\|\mathbf{b}\|_2 \equiv (\sum_i b_i^2)^{1/2}$. The regularized iterative reconstruction method yields an optimal solution to the equation system $\mathbf{A}\mathbf{x} = \mathbf{p}$ in the sense of both the data fidelity term and the regularization term being relatively small simultaneously. Namely, the solution achieves a good trade-off between data consistence/fidelity and the assumption contained in the regularization.

In the literature one may also find a constrained problem, written as

$$\min_{\mathbf{x}} R(\mathbf{x}), \quad \text{subject to} \quad \mathbf{A}\mathbf{x} = \mathbf{p}, \tag{3.28}$$

for the prior information-based reconstruction [20]. It aims to minimize the regularization term while enforcing the solution to be consistent with the measurements $\mathbf{p}$. However, this problem may have no solution at all considering the errors in experimental data. In this case a tolerance is required to obtain an approximate solution such that

$$\min_{\mathbf{x}} R(\mathbf{x}), \quad \text{subject to} \quad \|\mathbf{A}\mathbf{x} - \mathbf{p}\|_2^2 \leqslant \epsilon, \tag{3.29}$$

where ϵ is the tolerance of small value. Nevertheless, the constrained problem in (3.29) is more commonly reformulated to a variant form in (3.27) by introducing a Lagrangian multiplier literally [44].

The regularization $R(\mathbf{x})$ could be L1-norm that $R(\mathbf{x}) = \|\Phi\mathbf{x}\|_1$, or the square of L2-norm that $R(\mathbf{x}) = \|\Phi\mathbf{x}\|_2^2$. There the choices for Φ can be an identity matrix, a matrix approximating specific operator, or a domain transformation such as the discrete gradient transform used in total variation (TV) and the wavelet transform, in various applications. The L1-norm and L2-norm are defined as $\|\mathbf{b}\|_p \equiv (\sum_i |b_i|^p)^{1/p}$ with $p =$1 and 2 respectively. The underlying philosophy in the L1- and L2-norm regularization is that the image $\mathbf{x}$ has a sparse representation in the transformed domain denoted by Φ. If Φ is equal to an identity, it means that the image $\mathbf{x}$ itself is sparse in spatial domain. However, this is uncommon for an image since the gray values of its majority pixels are normally much larger than zero. Note that the Tikhonov regularization [117] has the form of L2-norm. It is only ever defined in terms of linear matrix operation and the corresponding minimization program has explicit solution. Thus application of the Tikhonov regularization may avoid of iterations. However, the L2-norm has smooth nature. Namely, penalty to large coefficients will lead to a larger reduction of the objective function compared to penalty to small coefficients. In contrary, the L1-norm is less sensitive to sharp edges which normally represent important features in images. Thus the L1-norm based regularization is generally used to induce a more sparse and unique solution to (3.27) while preserving the main features in the reconstructed image.

Another main problem of all regularization algorithms is the selection of the trade-off parameter λ and its influence on both the convergence and quality of the reconstructed image. A small λ keeps a good data fidelity, but leads to lots of noises and artifacts in the reconstructed image. In contrast, a large value of λ strengthens the regularization term to reduce the noises and artifacts, but a too strong regularization will remove the edges and some obvious features, resulting in an unrealistic reconstruction image.

The iterative reconstruction with regularization is to solve the optimization problem in Eq. 3.27, which is not exclusive for the reconstruction of X-ray computed tomography. In

many other signal or image processing applications, such as signal denoising [93], image restoration [94], reconstruction of ultrasound computed tomography (USCT) [95], and so on, the problems are often mathematically modeled into ill-posed equation systems which are solved by regularization method in order to integrate additional information. Due to the common usage of the mathematical model in the literature, various solvers proposed for the generic problem in (3.27) can be employed to achieve the reconstruction result.

Note that the tomographic problem arising in real-world tomography applications is not only in large scale but also involves dense matrix data which require large memory. This often precludes the use of second-order methods with quadratic convergence and motivates the study of simpler gradient-based algorithms for solving (3.27). Among the first-order based algorithms the dominant computational effort is a relatively cheap matrix-vector multiplication related to $\mathbf{A}$ and $\mathbf{A}^{\mathrm{T}}$ other than expensive computation involved in high-order derivative calculation.

Here it is necessary to do a brief study on the algorithms applicable in solving the problem (3.27). These algorithms can be separated into two groups. The first one includes certain classical optimization methods, such as steepest/gradient descent method, Newton's method, and conjugate gradient method. Apart from these algorithms, scientists have made more attempts in searching a robust and faster method to solve the optimization problem in (3.27). The advanced recent attempts include the iterative shrinkage-thresholding algorithm and its variants, variable splitting, and alternating minimization based algorithms. They comprise the second group of algorithms.

Classical optimization algorithms

A simple method to solve the optimization problem is the gradient descent method[20] which finds the minimum of the objective function by taking the descend directions proportional to the negative gradient of the objective function at the current point. Let $f(\mathbf{x})$ denote the objective function in (3.27). Thus the update of estimate in each iteration k is performed as

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - \alpha^{(k)} \nabla f(\mathbf{x}^{(k)}) \,, \tag{3.30}$$

where $\alpha^{(k)}$ denotes the step size in kth iteration determined by a line search method such as the back-tracking method [96]. The gradient descend is a first-order optimization algorithm only demanding the gradient information for updating descend directions. Higher order algorithms, such as Newton's method [97] and its variants, converges much faster towards to the local minimum than gradient descend. However, they might not be applicable since the higher order information such as Hessian matrix does not exist or involves expensive calculation.

Another option of the solver in solving (3.27) is the nonlinear conjugate gradient approach [98, 96]. This method is different from the gradient descend method in the way

that in each iteration (except the first iteration) it follows a conjugate direction which incorporates the direction in the former iteration and the descend gradient at current point. The conjugate direction $\mathbf{d}^{(\mathbf{k})}$ (for $k \geq 1$) is written as

$$\mathbf{d}^{(k)} = -\nabla f(\mathbf{x}^{(k)}) + \beta \mathbf{d}^{(k-1)} \qquad (3.31)$$

It still demands no more than the first order information. The first iteration is performed using the gradient descend algorithm, such as

$$\mathbf{d}^{(\mathbf{0})} = -\nabla f(\mathbf{x}^{(\mathbf{0})}) \,. \qquad (3.32)$$

Then we update the estimate along the conjugate direction $\mathbf{d}^{(\mathbf{k})}$ in the way that

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha^{(k)} \mathbf{d}^{(k)} \,. \qquad (3.33)$$

The nonlinear conjugate gradient algorithm works well if the objective function is approximately quadratic near the minimum. Otherwise, the algorithm becomes slower. The conjugate gradient method can follow the narrow valleys where the gradient descend method is not able to converge fast due to the criss-cross pattern it follows.

Advanced algorithms

Iterative shrinkage-thresholding (IST) algorithms [99, 100] are another popular class of methods for solving the linear inverse problems. They can be viewed as an extension of the classical gradient algorithm. In each iteration, they do not intent to search a descend direction for the whole objective function $f(\mathbf{x})$ in solving the problem (3.27) but apply the gradient descend method only for the data fidelity term followed by a shrinkage/soft-threshold step. The iteration of IST algorithm is specifically performed in the way that

$$\mathbf{x}^{(k+1)} = \Gamma_{\mathbf{L},\lambda'} \{ \mathbf{x}^{(k)} - 2\alpha^{(k)} \mathbf{A}^{\mathrm{T}}(\mathbf{A}\mathbf{x}^{(k)} - \mathbf{p}) \} \,, \qquad (3.34)$$

where $2\mathbf{A}^{\mathrm{T}}(\mathbf{A}\mathbf{x}^{(k)} - \mathbf{p})$ is exactly the derivative of $\|\mathbf{A}\mathbf{x}^{(k)} - \mathbf{p}\|_2^2$; $\alpha^{(k)}$ is an appropriate step size and $\Gamma_{\mathbf{L},\lambda'}$ is the shrinkage/soft thresholding operator applied on the coefficients in the transformed domain $\mathbf{L}$ in (3.27) with respect to the threshold parameter λ'. The soft thresholding operation on a coefficients vector $\mathbf{b} = \{b_i\}, i = 1, 2, \cdots, N$ is defined as

$$\Gamma_{\lambda'}(b_i) = \mathrm{sgn}(b_i) \cdot \max\{|b_i| - \lambda', 0\} \,. \qquad (3.35)$$

The parameter λ' used as a threshold in $\mathbf{L}$ domain is implicitly related to the regularization parameter λ in (3.27). Note that the operator $\Gamma_{\mathbf{L},\lambda'}$ in Eq. 3.34 also involves the forward and backward transform apart from the shrinkage/soft thresholding operation in Eq. 3.35.

IST algorithms are attractive due to the simplicity and adequate for solving large-scale problems even with dense matrix data. However, the IST algorithms are known to have such slow convergence rate as gradient descend method, which has been well analyzed in the literature [99, 101]. Recently accelerated versions for IST algorithms have been presented, such as TwIST [46], FISTA [26] and NESTA [47], all of them applying the idea that the next iteration is not only based on the previous one as in Eq. 3.34, but also on two or more previously computed iterations. The convergence analysis and the difference between these fast algorithms of IST have been elaborated in [26].

Note that the gradient descend method and nonlinear conjugate gradient method solving the regularized minimization problem in (3.27) require that the first-order derivative of the regularization term $R(\mathbf{x})$ must exist. IST algorithms separate the data fidelity term and regularization term using two steps to update the image respectively, where only the first-order derivative is involved for data fidelity term. Thus the IST algorithms avoid applying the derivative of $R(\mathbf{x})$.

3.4. Comparison

Now the reconstruction methods for computed tomography can be concluded by recalling Fig. 3.1 where they are categorized into different groups based on their properties. From the introduction of analytical and iterative reconstruction methods given in Sec. 3.2 and 3.3 their distinctions in terms of theoretical and practical implementation process can be easily understood. In the following we discuss the main advantages and disadvantages of these algorithms by comparing them in terms of reconstruction quality and computational requirements. FBP is taken as the representative method for analytical reconstruction.

Reconstruction quality

FBP is an algorithm with the reconstruction image quality highly depending on the number of measured projections. It implicitly assumes the coefficients in Fourier domain are zero between any two adjacent angles where projections are taken. The FBP method is able to produce good quality image for sufficient projection data. The amount of sufficient data was given in Sec. 3.2.3. However, a low number of projections leads to streak artifacts due to the insufficient data along the radial axis in Fourier domain. Apart from the number of projections, other factors like the errors in the measured projections and assumptions made in the processes of back-projection and filtering, also have important impact on the reconstruction image quality. With the current trend of reducing the projection data for practical purposes, the FBP reconstruction is no longer considered to be optimal.

The IR methods search the optimal solution in a numerical and iterative way. The main advantage of IR compared to the analytical FBP algorithm is its flexibility in incorporating more models and prior information. This enables the IR methods to produce higher quality image in certain cases where the FBP is helpless.

First, iterative reconstruction methods are easier to incorporate different models, representing physical, natural imaging processes or specific requirements, into the image reconstruction. In the forward imaging process, the X-ray beam is generally modeled as a straight line traveling through the object to the detector. The model can be adapted for different cases. For example, for a linear X-ray source the X-rays are modeled as parallel beams and for a point X-ray source the X-rays are divergent beams. The X-ray's interaction with the image pixels can also be modeled in different ways. IR methods can easily integrate these models into the reconstruction process without changing other main components of the algorithm. An accurate model will improve the reconstruction image quality.

Secondly, the iterative reconstruction methods are able to accommodate the prior information for searching the optimal solution in high dimensions. The prior information can be the statistical properties of noise in the measured data, the relationship of adjacent pixels in the reconstruction image, the sparseness of both the measured data and the reconstructed image. IR methods integrate these prior information using regularization. Due to the intrinsic difference in dealing with data between FBP and iterative reconstruction methods, reconstruction results from iterative methods may have different appearance from those using FBP reconstruction. The usage of regularization can greatly improve the image quality, especially if the projection data are uncompleted. The computed tomography with less projections is helpful in optimizing the CT system and accelerating the data acquisition process.

The introduction of regularization in the iterative reconstruction methods is helpful not only for reducing streak-line artifacts caused by the limited number of projections, but also for increasing the reconstruction robustness with respect to data inconsistencies caused by the following three types of errors [102].

- Error in the projection geometry

- Defect pixels in the measured projections

- Noise in the measured projections

The data inconsistencies of the projections lead to a large value for $\|\mathbf{A}\mathbf{x} - \mathbf{p}\|_2^2$. The regularization in iterative methods is assistant in controlling the relationship of adjacent pixels throughout the reconstruction procedure. It maintains the smoothness of the reconstruction image while preserving important details. A denoising approach of post-reconstruction filtering may also be used to produce smoothed reconstruction image in

order to reduce the impacts of data inconsistencies, but it can not well preserve important details in the image and tends to decrease the image accuracy and spatial resolution. The reconstruction results in [102] show that the iterative reconstruction method based on regularization is more robust to the inconsistencies in the projection data compared to the FBP reconstruction.

Iterative reconstruction methods have recently received much attention in computed tomography due to their superiority over the analytical FBP method mentioned above and motivated by requirements of reconstruction with limited number of projections in various scientific researches. However, one has to note the fact that the FBP-based reconstruction algorithms are still overwhelming in most tomographic imaging applications, especially the CT systems in business and clinical application. The usage of iterative reconstruction methods has been impeded by two points. One is the lack of computational efficiency. High computation load has always been a great challenge for iterative reconstruction. This will be discussed in the section below. The second point that constrains the application of iterative reconstruction methods is their relatively weak numerical stability. Due to the ill-posed property of the reconstruction problem of limited number of projections, there exists no unique solution. The iterative reconstruction produces a series of successive approximations with respect to the least square criterion to make the reconstruction image consistent with the projection data. Meanwhile the regularization leads the approximation approaching to a desirable point determined by the integrated prior information. The iterative path can be influenced by the initial guess, the stop condition, and the signal to noise ratio of projections. It probably converges to a local minimum with improper input parameters. Careful evaluation and parameter optimization are required before one incorporates the iterative reconstruction into routine CT systems in business and clinical practice.

Computational requirements

Both the FBP-based reconstruction and the iterative reconstruction are very computationally expensive given that real world data obtained from CT systems are relatively large. Memory management and arithmetic computation pose challenges in terms of computation time for real-time data analysis. Relatively speaking, iterative reconstruction is computationally more expensive compared with the FBP-based reconstruction considering the intrinsic difference in dealing with the projection data.

The FBP-based algorithms directly calculate the image in a single reconstruction process without any iterations. The reconstruction process consists of four successive steps as shown in Fig. 3.4. Among them the FT and IFT is not always necessary in practice considering that the filtering operation in Fourier domain is equivalent to convolution of a filter kernel in spatial domain with projections. Note that the back-projection step requires

a large amount of interpolation operations due to the coordinate mismatching between image pixels and projection bins in discrete domain. This step consumes the maximum time ($\sim 80\%$) of the FBP reconstruction [103]. Strategies for coping with this computational demand have been proposed, like heavily zero-padding the projection data which allows the selection of nearest point in back-projection [30]. The most effective and commonly applied method for accelerating the FBP reconstruction is parallel implementation of the FBP algorithm. It takes into account the computational independence of each projections in Fourier domain according to the Fourier slice theorem. High performance computing units, like GPU, can be easily employed for the parallel implementation of the four steps of FBP [104]. Consequently, the computing time required by FBP reconstruction is dramatically reduced. The FBP reconstruction can be optimized to a real-time reconstruction in many practical applications with assistant of high performance computing infrastructures.

In contrast, iterative reconstruction sequentially and iteratively constructs a series of approximations to unknowns $\mathbf{x}$. Each iteration consists of three major steps as shown in Fig. 3.6: creating artificial projection data by forward projection, computing correction term by comparing artificial and measured projection data, and updating the object by back-projecting the correction term. Both the forward and backward projection are computationally demanding operations due to multiplications of large matrices required. The iterative process is completed when either a fixed number of iterations is reached or the update of current image approximation is small enough. It normally contains hundreds of iterations in order to obtain an accurate reconstruction image. This large number of iterations must be implemented sequentially in terms of algorithm, thus further making the iterative reconstruction computationally demanding.

Due to the high requirements on the computing power, iterative reconstruction methods are not widely used at the beginning, when the computer technology was far from satisfying the requirements. As the development of advanced algorithms and modern computer technology, the iterative reconstruction on general computer becomes practical. Several strategies have been proposed to accelerate the iterative reconstruction. One searches for iterative algorithms with fast convergence to decrease the number of iterations required. Another strategy uses high performance computing units like GPU exclusively for specific time consuming operations such as forward and backward projection. For further improving the reconstruction speed, devices of high computing performance, such as GPU and CPU cluster [105, 24], are commonly applied for parallel reconstruction in data level, such as stacking all slice reconstructions to a volumetric image. Software and hardware methods are being investigated to accelerate iterative reconstruction. However, one should note that for a reconstruction problem with too large scale, such as a real 3D image reconstruction, it seems to still be a difficult task even with modern devices of high computing performance. With further developments in computational technology, iterative reconstruction may be incorporated into more practical CT systems in the future.

3.5. Summary

In this chapter the reconstruction methods for computed tomography are summarized. They are generally divided into two groups: analytical reconstruction methods and iterative reconstruction methods. The latter one is further categorized to the methods with and without regularization depending on whether additional information is integrated or not. FBP, as the typical analytical method, is able to perform the reconstruction process in real time, thus used by most of CT systems. However, streak artifacts becomes serious if the number of projections is not sufficient. Iterative reconstruction methods are much easier to incorporate forward models, statistical models of noise, and prior information. They are more robust in the cases of data inconsistency and insufficiency, reducing the noise and artifacts. As the focus of this thesis is on addressing the CT reconstruction problem with limited projection data, we mainly resort to the iterative reconstruction methods with regularization in order to combine the prior information (sparsity) and constrain the streak artifacts, while these ART-like or EM-like methods without regularization are powerless in this direction.

4. CT reconstruction with limited number of projections

There are several situations in which the CT reconstruction needs to be performed with limited number of projections. The first situation includes the clinical diagnosis, medical and biological research where the X-ray radiation dose is an issue of key importance. Reducing the number of projections can effectively constrain the radiation dose. Secondly, the *in vivo* investigation of internal, moving processes benefit from highest volume-acquisition rates, also resulting in low number of projections, e.g. the heart. Another case leading to imperfect projection data is the restriction in the scanning configuration where only the projections in a tilt angle of $\pm\theta$ ($\theta < 90°$) are available.

In the following of this chapter the compressive sampling (CS) theory and the CT reconstruction method for limited projection data based on this theory will be discussed. This chapter aims to solve the optimization problem with TV regularization using the nonlinear conjugate gradient (CG) method. Consequently, a method named CGTV is simply developed in this chapter which demands only the first derivative information of the objective function.

4.1. Compressive sampling in CT

The conventional data processing technique is based on the Nyquist-Shannon theory which samples a signal at the frequency at least twice the maximum frequency present in the signal. Recently a new theorem called compressive sampling (CS) [106], also known as compressive sensing, or sparse sampling, has been studied a lot for data acquisition and recovery. It claims that much fewer samples than that required by Nyquist-Shannon theory can recover the original signals which are actually sparse. The CS theory highly relies on an assumption of existence of a sparse representation. This is true for many natural signals, such as the acoustic waves and images representing the distribution of particular physical measures. The signal is not necessary to be sparse in the time or spatial domain, but should have a concise representation in a different transformed domain.

4.1.1. Sparsity

In computed tomography two or three dimensional images are involved. Let f denote a function of an image defined in the spatial domain. Using an orthogonal basis denoted by $\Phi = [\phi_1, \phi_2, \ldots, \phi_n]$, it can be expressed as:

$$f = \sum_{i=1}^{n} a_i \phi_i \,. \tag{4.1}$$

Consequently, the function f is represented by new coefficients a_i $(i = 1, 2, \cdots, n)$ in the new domain Φ. The image function f is defined as S-sparse if its representation a_i has only S coefficients that are obviously larger than zero.

To find a sparse representation in a proper domain is a key problem for the CS-based data processing. For instance, two image transform methods commonly used in digital image processing are discrete gradient transform (DGT) and discrete wavelet transform (DWT). In Fig. 4.1, we show examples of DGT and DWT by using two images (the Shepp-Logan phantom and Barbara image). The expressions for DGT in Fig. 4.1b will be given in next section. The wavelet transform used in Fig. 4.1c is a Haar wavelet, where the original image is filtered by high-pass filter, producing three large images corresponding to the horizontal, vertical, and diagonal coefficients respectively. The high-pass filtered images describe the local changes of brightness in the original image. An approximation image, shown in the upper-left, is obtained by low-pass filtering and downscaling the original image. The high-pass and low-pass filtering operation can be performed multiple times to produce the coefficients in different levels. In Fig. 4.1c two-level wavelet transform result is shown. As we can see, both the DGT and DWT coefficients are sparse. Namely, a large fraction of these coefficients are equal or close to zero. Only a small fraction of these coefficients are much larger than zero. They tend to be clustered around edges in the image and take the main energy of the image at high frequency. For image reconstruction one can get rid of these small coefficients without too much loss. Sparsity is a fundamental model allowing efficient image processing, such as image compression, segmentation, reconstruction, and so on.

4.1.2. Undersampling and sparsity-based reconstruction

Now let $\mathbf{x}$ denote the discrete spatial representation of f. It has sparse representation in domain Φ. The transformation of the image $\mathbf{x}$ is denoted by

$$\mathbf{x}' = \mathbf{\Phi}(\mathbf{x}) \,. \tag{4.2}$$

<div style="text-align:center">(a)   (b)   (c)</div>
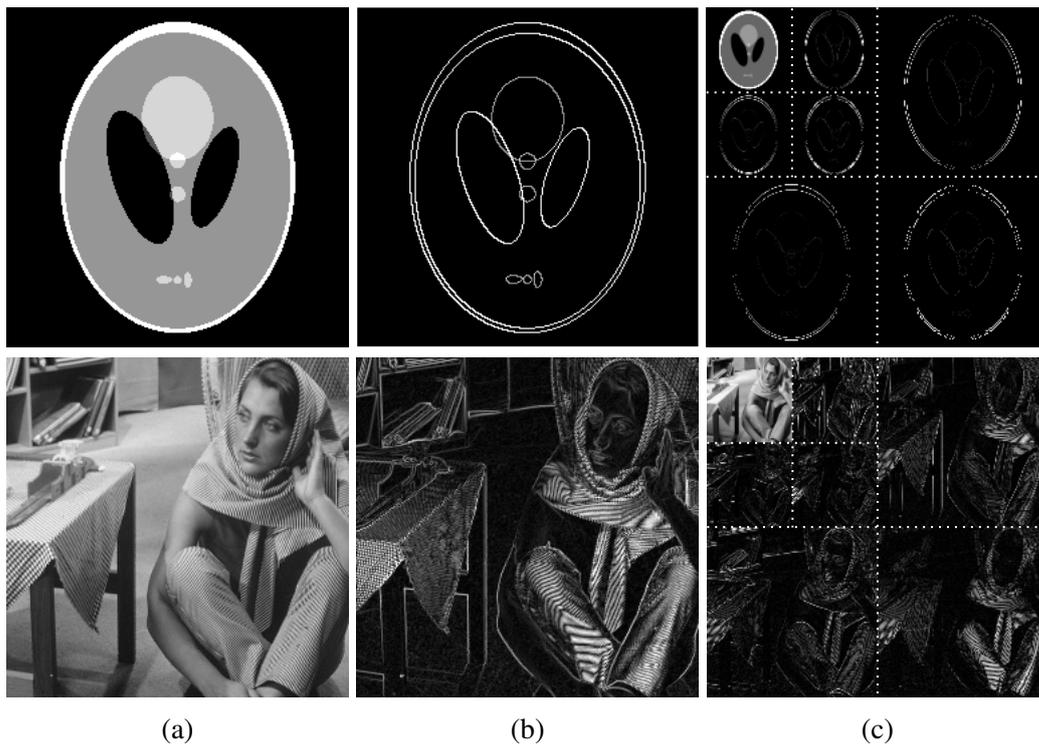
Fig. 4.1. Image representations in different domains: (a) the spatial representation of the Shepp-Logan phantom and the Barbara image and their sparse representations in (b) the discrete gradient transform (DGT) domain and (c) the discrete wavelet transform (DWT) domain.

Another basis $\boldsymbol{\Psi}$ is used for sensing/measuring the image $\mathbf{x}$. The measurements are

$$\mathbf{y} = \boldsymbol{\Psi}\mathbf{x}. \tag{4.3}$$

Ideally, it is expected to obtain sufficient measurements of $\mathbf{x}$, but one is actually able to observe a subset of them. Namely, the size of $\mathbf{y}$ is smaller than that of $\mathbf{x}$. This undersampling of $\mathbf{x}$ leads to an ill-posed inverse problem which can not find a unique solution. However, considering its sparse representation in $\boldsymbol{\Phi}$, one can recover the image $\mathbf{x}$ by minimizing the L1-norm of $\mathbf{x}'$

$$\min \|\mathbf{x}'\|_1 \qquad \text{subject to} \quad \mathbf{y} = \boldsymbol{\Psi}\mathbf{x}. \tag{4.4}$$

That is, among all solutions to $\mathbf{y} = \boldsymbol{\Psi}\mathbf{x}$, we pick the one whose representation in $\boldsymbol{\Phi}$ has minimal L1-norm. Here the L1-norm minimization is the way used to reconstruct the sparse solution.

The sparsity-based reconstruction can be simply transformed into the version used for the algebraic tomographic reconstruction of an object image. There the basis $\boldsymbol{\Psi}$ for measuring image is replaced by a system matrix $\mathbf{A}$. The measurements are denoted $\mathbf{p}$. That is $\boldsymbol{\Psi} \rightarrow \mathbf{A}$, $\mathbf{y} \rightarrow \mathbf{p}$. Thus the tomographic data are reconstructed through the L1-norm minimization, written as

$$\min_{\mathbf{x}} \|\boldsymbol{\Phi}(\mathbf{x})\|_1 \qquad \text{subject to} \quad \mathbf{A}\mathbf{x} = \mathbf{p}. \tag{4.5}$$

The regularization is defined as the L1-norm of the sparse representation of $\mathbf{x}$

$$R(\mathbf{x}) = \|\boldsymbol{\Phi}(\mathbf{x})\|_1. \tag{4.6}$$

The sparse transformation $\boldsymbol{\Phi}(\mathbf{x})$ can be the discrete gradient transform (DGT) or the discrete wavelet transform (DWT). Considering that the tomographic projections are often corrupted by noise during acquisition and transmission, the constraint in (4.5) should not exactly be an equation system but is changed to the form of inequations like

$$\min_{\mathbf{x}} \|\boldsymbol{\Phi}(\mathbf{x})\|_1 \qquad \text{subject to} \quad \|\mathbf{A}\mathbf{x} - \mathbf{p}\|_2^2 \le \epsilon. \tag{4.7}$$

ϵ is a small positive value. The existing of errors in the measured projections is another reason for the ill-posedness of the inverse problem, which is also expected to be solved by integrating the L1-norm of a sparse representation of the image. More often, the reconstruction problem is solved via the following unconstrained minimization

$$\min_{\mathbf{x}}\{\|\boldsymbol{\Phi}(\mathbf{x})\|_1 + \frac{\alpha}{2}\|\mathbf{A}\mathbf{x} - \mathbf{p}\|_2^2\} \tag{4.8}$$

by applying the multiplier method to (4.5). In this case, α represents the Lagrange multiplier associated with the corresponding constraint in (4.5). Note that (4.8) can be equivalently expressed as

$$\min_{\mathbf{x}}\{\|\mathbf{A}\mathbf{x} - \mathbf{p}\|_2^2 + \lambda\|\mathbf{\Phi}(\mathbf{x})\|_1\}\,, \tag{4.9}$$

that is the same as (3.27).

An important concept of CS for data recovery for the linear system $\mathbf{A}\mathbf{x} = \mathbf{p}$ is the so-called restricted isometry property (RIP) [106]. It requires that all subsets of S columns in the system matrix $\mathbf{A}$ are nearly orthogonal. This condition is not fulfilled in CT considering that the projections are obtained by the regularly arranged detectors and equiangularly distributed measurements. Strictly speaking, CS is not applicable to CT imaging since the RIP concept is too restrictive. The number of projections required for an accurate CT reconstruction can not be exactly calculated. Despite this fact, the CS theory is still potentially useful in a practical sense for CT reconstruction by using Eq. 4.6 in the optimization problem (4.9).

4.2. Total variation-based reconstruction

Pursuing the CS idea of sparsity-base reconstruction, a base change called discrete gradient transform (DGT) can lead to a sparse expansion of most 2D or 3D images. The totality of associated expansion coefficients expresses the so-called total variation (TV) of the image. The iterative reconstruction methods with TV regularization have been studied for limited projection data problems [48, 20, 49].

For later use, it is convenient to represent the object in terms of a 2D matrix $\mathbf{X} \in \mathbb{R}^{n \times n}$ (n pixels in each of the two dimensions) where each entry is associated with the local value of the attenuation index. Thus, vector $\mathbf{x}$ in (4.9) can be composed out of the elements $\mathbf{X}_{i,j}$ ($i, j = 1, \cdots, n$) by column $\mathbf{X}_{i,1}$ being stacked on top of column $\mathbf{X}_{i,2}$ and so forth. Here the dimensionality of vector $\mathbf{x}$ is given as $N = n^2$ (total pixel number in the to-be-reconstructed slice).

The system matrix $\mathbf{A} \in \mathbb{R}^{M \times N}$ in (4.9) possesses M rows and $\mathbf{p} \in \mathbb{R}^{M \times 1}$. For exact reconstruction one would have to impose $M = N$ but usually one has $M < N$ in the case of limited projection data such that a regularization for a reasonable solution to the reconstruction problem in the sense of (4.9) is required. The function to be minimised in (4.9) consists of two terms: the data fidelity term, $F(x) = \|\mathbf{A}\mathbf{x} - \mathbf{p}\|_2^2$, and the total variation term (regularization), $R(\mathbf{x}) = \|\mathbf{x}\|_{\text{TV}}$. To define it, we consider the L1-norm or the L2-norm of the gradient field obtained by the so-called discrete gradient transform (DGT). For a given pixel (i, j) one has

$$\text{DGT}_{\text{l1},ij}(\mathbf{X}) \equiv |\triangle_{i,j}^h \mathbf{X}| + |\triangle_{i,j}^v \mathbf{X}| \tag{4.10}$$

or

$$\mathrm{DGT}_{l2,ij}(\mathbf{X}) \equiv [(\triangle_{i,j}^{h}\mathbf{X})^2 + (\triangle_{i,j}^{v}\mathbf{X})^2]^{1/2}\,, \tag{4.11}$$

where the difference operations $\triangle_{i,j}^{h}\mathbf{X}$ (h for horizontal) and $\triangle_{i,j}^{v}\mathbf{X}$ (v for vertical) on matrix $\mathbf{X}$ are defined as

$$\triangle_{i,j}^{h}\mathbf{X} \equiv \mathbf{X}_{i,j} - \mathbf{X}_{i-1,j} \tag{4.12}$$

and

$$\triangle_{i,j}^{v}\mathbf{X} \equiv \mathbf{X}_{i,j} - \mathbf{X}_{i,j-1}\,, \tag{4.13}$$

respectively. Thus, total variations, defined in terms of the L1-norm or L2-norm of the gradient domain, see Eqs. (4.10) and (4.11), read

$$\|\mathbf{x}\|_{l1,\mathrm{TV}} \equiv \sum_{i,j} \mathrm{DGT}_{l1,ij}(\mathbf{X}) \tag{4.14}$$

or

$$\|\mathbf{x}\|_{l2,\mathrm{TV}} \equiv \sum_{i,j} \mathrm{DGT}_{l2,ij}(\mathbf{X})\,. \tag{4.15}$$

The usage of TV as a regularization in (4.9) implicitly assumes that the image is piece-wise smooth. Namely, in the image to reconstruct certain physically adjacent elements have (approximately) equal values with high probability. It intends to eliminate the artifacts and noises in the reconstruction but preserve its main features such as edges. The drawback of TV-base regularization method is that the assumption of the image being piece-wise smooth may cause blocky effects with a loss of fine structures in the reconstruction especially when the trade-off parameter λ in (4.9) is set to be overwhelming. To overcome this issue, some work on variant TVs, such as nonlocal total variation (NLTV) [107] and adaptive-weighted total variation (AwTV) [108], can be found in the literature.

4.3. CGTV solver

For the reconstruction problem in (4.9) the total variation (TV) is selected as the regularization so that $R(\mathbf{x}) = \|\mathbf{x}\|_{\mathrm{TV}}$. Consequently, the optimization in (4.9) can be rewritten as

$$\min_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{p}\|_2^2 + \lambda \|\mathbf{x}\|_{\mathrm{TV}}\,. \tag{4.16}$$

Two forms of TV have been defined in Sec. 4.2, $\|\mathbf{x}\|_{l1,\mathrm{TV}}$ and $\|\mathbf{x}\|_{l2,\mathrm{TV}}$, based on the L1-norm and L2-norm respectively in the gradient field obtained by the discrete gradient transform (DGT). The objective function to be minimized in (4.16) consists of two terms: the data fidelity term, $F(x) = \|\mathbf{A}\mathbf{x} - \mathbf{p}\|_2^2$, and the regularization term (TV),

$R(\mathbf{x}) = \|\mathbf{x}\|_{\mathrm{TV}}$. To distinguish different regularization, we let $R(\mathbf{x}) \rightarrow T(\mathbf{x})$ for TV regularization. Thus the objective function $f \equiv F(\mathbf{x}) + \lambda T(\mathbf{x})$.

Now let us turn to technical points in addressing the minimization problem in (4.16). To do this, the nonlinear conjugate gradient method [98] is employed. There are accelerated minimization schemes such as TwIST [46], UPN [48], FISTA [26] and ADMM [25]. Their common feature is that at least one parameter controlling the trade-off between data fidelity and regularization is determined empirically. It is true that, compared to these fast algorithms, the conjugate gradient method possesses higher iteration complexity. However, as such it exhibits fairly general, stable convergence properties. This is why the nonlinear conjugate gradient method is simply selected here. The resulting algorithm of solving 4.16 based on the nonlinear conjugate gradient method is called CGTV solver.

The nonlinear conjugate gradient method minimizing the objective function $f(x)$ requires the computation of the gradients of F and T in an n^2-dimensional vector space. The gradient of $F(\mathbf{x})$ is simply given as

$$\nabla F(\mathbf{x}) = \nabla(\|\mathbf{A}\mathbf{x} - \mathbf{p}\|_2^2) = 2\mathbf{A}^{\mathrm{T}}(\mathbf{A}\mathbf{x} - \mathbf{p}). \tag{4.17}$$

However, the expressions of TV in Eqs. 4.14 and 4.15 are not differentiable if $\mathbf{X}_{i,j} = \mathbf{X}_{i-1,j}$ and $\mathbf{X}_{i,j} = \mathbf{X}_{i,j-1}$. This is overcome by introducing a mild modification of the modulus definition as [109]

$$|z| \approx (z^2 + \epsilon)^{1/2}, \tag{4.18}$$

where ϵ is a smoothing parameter much smaller than typical values of z. Using this approximation, the gradient of (4.18) can be written as $\nabla|z| \approx z/(z^2 + \epsilon)^{1/2}$. Therefore, the partial derivative of Eqs. 4.14 and 4.15 with respect to $\mathbf{X}_{i,j}$ are obtained as follows:

$$\frac{\partial \|\mathbf{x}\|_{\mathrm{l1,TV}}}{\partial \mathbf{X}_{i,j}} = \frac{\mathbf{X}_{i,j} - \mathbf{X}_{i-1,j}}{[(\mathbf{X}_{i,j} - \mathbf{X}_{i-1,j})^2 + \epsilon]^{1/2}} + \frac{\mathbf{X}_{i,j} - \mathbf{X}_{i,j-1}}{[(\mathbf{X}_{i,j} - \mathbf{X}_{i,j-1})^2 + \epsilon]^{1/2}}$$
$$+ \frac{\mathbf{X}_{i,j} - \mathbf{X}_{i+1,j}}{[(\mathbf{X}_{i+1,j} - \mathbf{X}_{i,j})^2 + \epsilon]^{1/2}} + \frac{\mathbf{X}_{i,j} - \mathbf{X}_{i,j+1}}{[(\mathbf{X}_{i,j+1} - \mathbf{X}_{i,j})^2 + \epsilon]^{1/2}} \tag{4.19}$$

and

$$\frac{\partial \|\mathbf{x}\|_{\mathrm{l2,TV}}}{\partial \mathbf{X}_{i,j}} = \frac{2\mathbf{X}_{i,j} - \mathbf{X}_{i-1,j} - \mathbf{X}_{i,j-1}}{[(\mathbf{X}_{i,j} - \mathbf{X}_{i-1,j})^2 + (\mathbf{X}_{i,j} - \mathbf{X}_{i,j-1})^2 + \epsilon]^{1/2}}$$
$$+ \frac{\mathbf{X}_{i,j} - \mathbf{X}_{i+1,j}}{[(\mathbf{X}_{i+1,j} - \mathbf{X}_{i,j})^2 + (\mathbf{X}_{i+1,j} - \mathbf{X}_{i+1,j-1})^2 + \epsilon]^{1/2}}$$
$$+ \frac{\mathbf{X}_{i,j} - \mathbf{X}_{i,j+1}}{[(\mathbf{X}_{i,j+1} - \mathbf{X}_{i-1,j+1})^2 + (\mathbf{X}_{i,j+1} - \mathbf{X}_{i,j})^2 + \epsilon]^{1/2}}. \tag{4.20}$$

The gradient of objective function $f(\mathbf{x})$ is now calculated by $\nabla f(\mathbf{x}) = \nabla F(\mathbf{x}) + \lambda \nabla T(\mathbf{x})$ using $\nabla F(\mathbf{x})$ in Eq. (4.17) and $\nabla T(\mathbf{x})$ in Eq. 4.19 or Eq. 4.20.

Algorithm 1 below shows how the CGTV solver is implemented [110]. In this algorithm, the parameter $\beta^{(k)}$ controls the update of conjugate direction. In the implementation here the Dai and Yuan formula for $\beta^{(k)}$ proposed in [111] is applied, showing robust and fast convergence property which is absent for other formulas such as Hestenes-Stiefel, Fletcher-Reeves, Polak-Ribière, etc., listed in [98].

Algorithm 1 TV-based conjugate gradient method (CGTV)

Input: set ϵ to be a positive, small value (typically $\epsilon = 10^{-6}$); take $\mathbf{x}^0$ as the initial guess; prescribe parameter value for λ; maximum iteration number K.

Output: estimate $\mathbf{x}$

1: $k := 0$
2: **while** $k \leq K$ **do**
3:     calculate gradient direction $\nabla f(\mathbf{x}^{(k)}) = \nabla F(\mathbf{x}^{(k)}) + \lambda \nabla T(\mathbf{x}^{(k)})$
4:     set Dai and Yuan formula [111] $\beta_{\mathrm{DY}}^{(k)} = \|\nabla f(\mathbf{x}^{(k)})\|_2^2 / [(\mathbf{d}^{(k-1)})^{\mathrm{T}}(\nabla f(\mathbf{x}^{(k)}) - \nabla f(\mathbf{x}^{(k-1)}))]$ for $k \geq 1$, $\beta_{\mathrm{DY}}^{(0)} = 0$
5:     update the conjugate direction $\mathbf{d}^{(k)} = -\nabla f(\mathbf{x}^{(k)}) + \beta_{\mathrm{DY}}^{(k)} \mathbf{d}^{(k-1)}$
6:     perform back-tracking line search algorithm to find step size $\alpha^{(k)}$
7:     update the object estimate $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha^{(k)} \mathbf{d}^{(k)}$
8:     **if** $\|\alpha^{(k)} \mathbf{d}^{(k)}\| / \|\mathbf{x}^{(k)}\| < \epsilon$ **then**
9:         **return** estimate $\mathbf{x}^{(k+1)}$
10:    **end if**
11:    $k = k + 1$
12: **end while**
13: **return** estimate $\mathbf{x}^{(K)}$

In practice, no essential differences arise when applying either of the definitions of total variation in Eq. (4.14) or Eq. (4.15). However, definition of (4.15) does not depend on the choice of Cartesian coordinates, and therefore it will be exclusively used throughout the practical implementation of Algorithm 1. Thus, notationally, we let $\|\mathbf{x}\|_{\mathrm{l2,TV}} \rightarrow \|\mathbf{x}\|_{\mathrm{TV}}$ for all reconstructions in this thesis.

4.4. Reconstruction with the CGTV solver

Due to the fact that the CGTV solver applies the concept of compressive sampling theory, it is able to reconstruct a high quality image from limited number of projections compared to the traditional FBP reconstruction. To evaluate the reconstruction quality of the CGTV solver, two data sets with the reconstruction problem of limited number of projections are examined here. One represents an insect imaged by propagation-based X-ray computed tomography at ANKA. The imaging is a mixture of phase and absorption intensity

contrast in order to enhance issues' border in biological body. This biological data set will be used to examine the CGTV reconstruction if the number of projections decrease to a small value. The second data set is acquired from electron tomography imaging of nanostructured materials for visualizing and analyzing 3D material structures in nanometer resolution. In this data set, not only just a small number of projections are available, but also an angular wedge is missing in projections. The missing wedge will further give rise to artifacts in conventional reconstruction. The reconstruction method based on compressive sampling theory is expected to decrease the artifacts caused by both limited number of projections and missing wedge.

4.4.1. Reconstruction of X-ray tomography

The X-ray tomography data set acquired at tomography beamline of ANKA is an intensively sampled result (projections) in 1500 angles between 0° and 180°. These projections are exported as a series of TIF images of size 1024×1024. The sinogram of a 2D slice is 1024×1500. In order to apply the regularized iterative reconstruction method to the limited projection data, only a subset of these 1500 projections is required. Thus the angular resolution of the dataset is reduced by equi-angularly extracting 60 projections from 1500 to compose a new sub-dataset. Consequently, the size of sinogram decreases to 1024×60, taking only 4% of the original data. In Table 4.1 the major parameters related to the original data set and sub-dataset are shown. The reconstruction image is set to be 512×512 here, which is a compromise between the image accuracy and requirements in computation resources and computing time.

Table 4.1. Major parameters of experimental dataset

	radiograph	number of projections	rotation center	sinogram
original data	1024×1024	1500	599.3	1024×1500
sub-dataset	1024×1024	60	599.3	1024×60

Figure 4.2 shows the reconstruction results of a slice of the X-ray tomography sub-dataset. There Fig. 4.2a is a referenced reconstruction image obtained by FBP reconstruction using the original dataset, where minor artifacts exist since sufficient projections (1500) are available for a high resolution reconstruction. The other two images in Figs. 4.2b and 4.2c are reconstructed using the sub-dataset (60 projections) by FBP method and CGTV solver (regularized iterative method) respectively. The FBP method achieves an image with serious streakline artifacts, which, however, are greatly reduced in the image obtained by the regularized iterative reconstruction method. The reconstruction quality is visually close to the reference image even though the volume of sub-dataset is only 4% of that

<center>(a)   (b)   (c)</center>
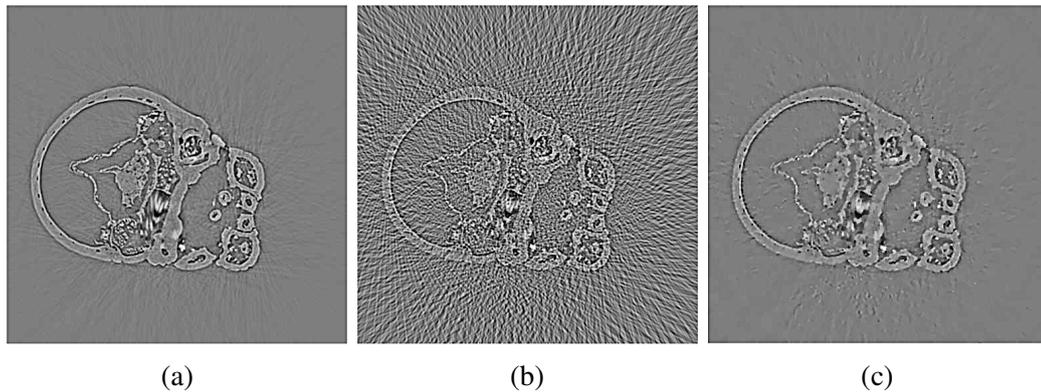
Fig. 4.2. Reconstruction results of a slice of X-ray tomography data: (a) reference image reconstructed using the FBP method from the original data set (1500 projections); (b) reconstructed image using the FBP method from the sub-dataset (60 projections); (c) reconstructed image using the regularized iterative method (CGTV) from the sub-dataset (60 projections).

of the original dataset. Thus the regularized iterative method (CGTV) is proved to be a better way in addressing the artifacts error in FBP reconstruction in the case of limited number of projections in computed tomography.

4.4.2. Reconstruction of electron tomography

The electron tomography data have the same mathematical model for forward imaging and reconstruction with the X-ray tomography data even though their imaging methods are based on different physical principles. The electron tomography data are from the imaging of a sample comprised of two different materials: metal particles and a support material. For more details about the formulation of the sample and materials see [112]. The data contain 75 projections over a tilt-range of $\pm 74°$ in an increment of $2°$. Projections are missing in an angular wedge of $32°$. The sinogram of a slice is 600×75. The reconstruction image is set to be 600×600. The aim of this electron tomography data is to automatically segment the metal particles from the reconstruction. Let us see how the regularized iterative method behaves in addressing artifacts in the reconstruction caused by limited number of projections and missing angular wedge.

Reconstruction results of a slice by using FBP and regularized iterative method (CGTV) are displayed in Fig. 4.3. All reconstructions are normalized to [0,1] in gray scale. The sample is comprised of metal particles and support material showing different values in the reconstruction image. The same material should have a homogeneous value. Fig. 4.3a and 4.3b respectively show the FBP and CGTV reconstruction by using 75 projection in $\pm 74°$. From visual judgment the FBP reconstruction obtain a bad quality image with

streakline artifacts and wedge artifacts as well. In contrast, the CGTV reconstruction well reduces these artifacts in 4.3a and produces roughly homogeneous areas for the two different materials. For further comparison the CGTV reconstruction using 180 projections in the angular range of $\pm 90°$ in an increment of $1°$ is given in Fig. 4.3c, which shows a reconstruction image close to the image representing the real object. Due to the lack of projections in an angular wedge, the upper and lower edges of the support material in 4.3b are not well reconstructed.


(a)

(b)

(c)
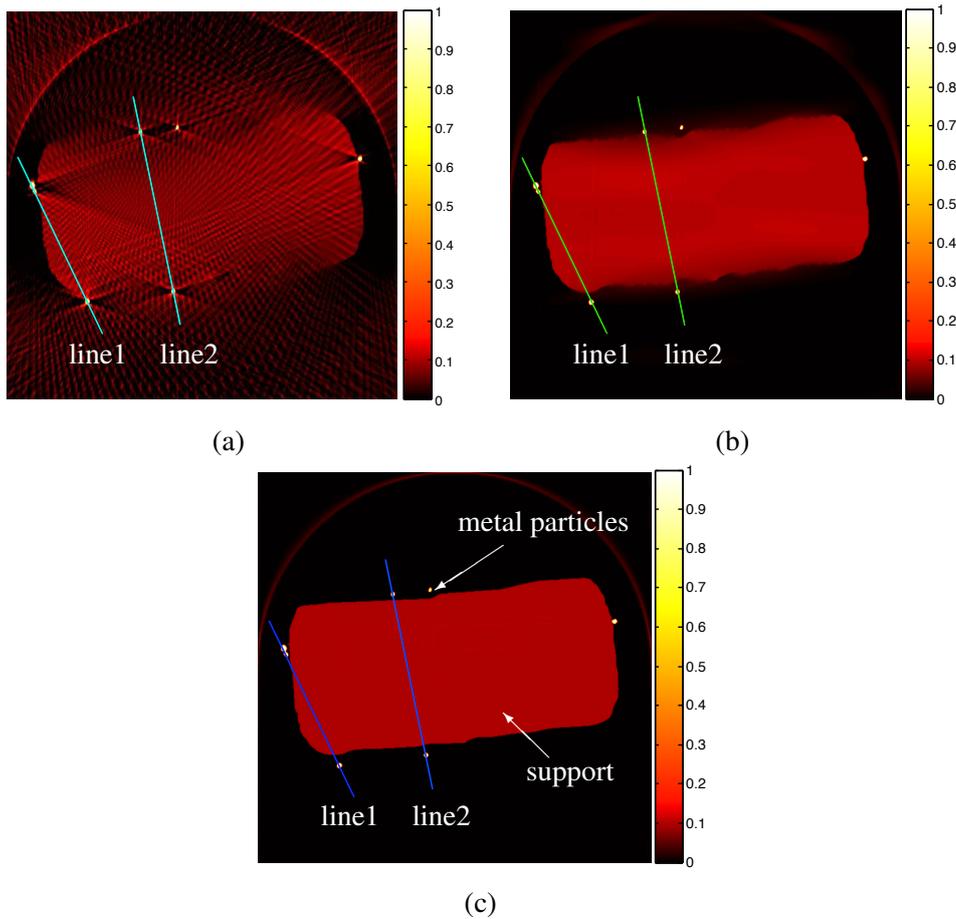
Fig. 4.3. Reconstruction results of a slice of electron tomography data: (a) FBP reconstruction by using 75 projections in $\pm 74°$; (b) CGTV reconstruction by using 75 projections in $\pm 74°$; (c) CGTV reconstruction by using 180 projections in angular range of $\pm 90°$.

To further illustrate the properties of the FBP and CGTV reconstruction in the case of low number of projections and missing angular wedge, two line cuts through the slice of metal particles and support material are presented in Fig. 4.4. The positions of the two lines, 'line1' and 'line2', are displayed in Fig. 4.3. Gray values along these two lines in the reconstructions of 4.3a (marked in cyan), 4.3b (marked in green), and 4.3c (marked

in blue) are plotted in Fig. 4.4a and 4.4b respectively. Artifacts generated by FBP reconstruction result in great errors in the reconstruction image destroying the homogeneity of material areas. In contrast, the CGTV reconstruction preserves better the edges and homogeneity of material areas. Uncertainty of the upper and lower edges of support material in Fig. 4.3b can be also seen from the line cut in Fig. 4.4b. Small actual distances between the particles and support in this slice result in undershoots on both sides of the support in the two line cuts. These undershoots disappear in the FBP reconstruction of 4.3a along 'line1' and 'line2', and also in the CGTV reconstruction of 4.3b along 'line2'. The edges of particles can be easier identified in the CGTV reconstruction than the FBP reconstruction, which will be helpful in the next step of automatically segmenting the metal particles more accurately. To sum up, the regularization used in CGTV reconstruction method successfully reduces the artifacts appearing in FBP reconstruction caused by low number of projections and missing angular wedge.


(a) line1   (b) line2
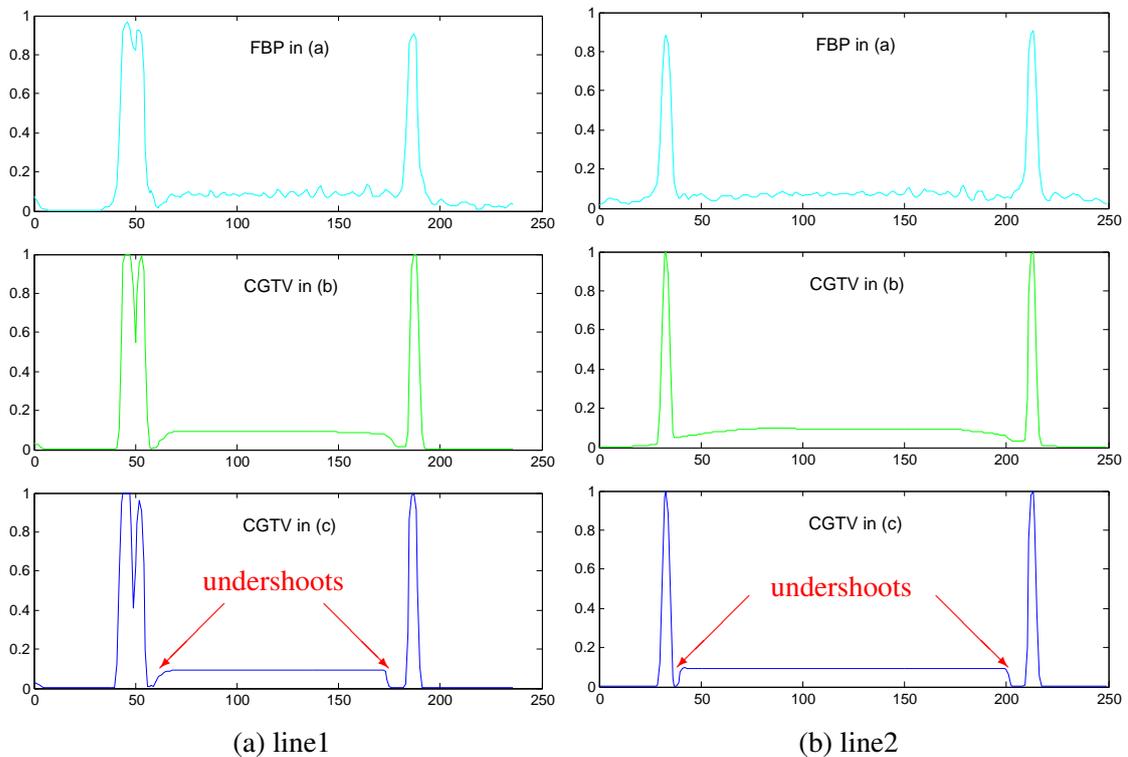
Fig. 4.4. Gray values (normalized to [0,1]) along the two lines through the reconstructions, (a), (b), and (c) in Fig. 4.3: (a) for 'line1' and (b) for 'line2'. Horizontal axes represent the distance in pixel.

The same results can be obtained for more real world datasets [113, 114]. Note that the reconstruction images by CGTV shown in Fig. 4.2c and 4.3b involved a manual determination process for the trade-off parameter λ by visual judgment. The manually determined λ is equal to 0.1 for the X-ray tomography data and 1 for the electron tomography data.

Despite the high quality reconstruction by the TV-based regularization method compared to FBP, it has a drawback that it may cause blocky effect with a loss of fine structures especially when the trade-off parameter is set to be overwhelming. This can be slightly seen from the reconstruction results in Fig. 4.2c. The sparsity assumption made behind the regularization requires the original object being sufficiently sparse, otherwise any elements, such as fine structures making the object unsparse, could be removed by smoothing. This contributes to blocky effect in the reconstruction image.

4.5. Summary

The compressive sampling theory implies the possibility of reconstructing a precise image even from uncompleted tomographic data. The sparsity of the image can be used as prior information and incorporated into the regularization term in the optimization problem (4.9). The regularization term iteratively directs the solution to a reasonable one in the sense of satisfying the requirements of the prior knowledge. The regularization could be L1-norm, L2-norm, total variation (TV), or other norms based on sparse transformation such as the wavelet transform.

In this chapter, the total variation is selected as the regularization in order to employ the sparsity of the image in the gradient transform domain. A CGTV method is developed to solve the minimization problem for the tomographic reconstruction by using the nonlinear conjugate gradient method. The usage of TV regularization implicitly assumes the image is piece-wise smooth, able to eliminate the artifacts and noise in the reconstruction while preserving the main edges. Reconstruction results show that the TV regularization when applied to the biological X-ray tomography data and electron tomography data for material analysis allows for reliable sparsity-driven CT imaging with significantly reduced number of projections and even a missing angular wedge. It is able to not only improve the reconstruction quality but also reduce the data acquisition time compared to the conventional method. However, the regularization term should not be over-weighted to avoid the loss of fine structures and constrain the appearance of blocky areas. A proper value determined for the trade-off parameter λ is of great importance for the reconstruction quality.

5. Optimized CGTV for automated reconstruction

The trade-off parameter λ existing in most methods of solving the optimization problem (4.16) has an important influence on the quality of reconstruction image. In some methods, such as TVAL3 [49], ASD-POCS [38] and ADDM [25], this parameter is displaced by additional parameters which can equivalently control the balance between data consistency and reduction of noise and artifacts. Appropriately selected values of this parameter are of great importance. Since the proper values of these parameters vary for different datasets, they are often manually determined in practical implementation, which generally requires expert's experience in the field of CT reconstruction. The manual parameter determination is a very cumbersome process and thus restrains the construction of an automated workflow for the tomographic data processing and analysis.

In this chapter the goal is to keep the single parameter λ in (4.16) and solve the optimization problem with TV regularization using the nonlinear conjugate gradient (CG) method. The parameter λ in the CGTV method is determined with the assistance of the discrete L-curve method for the purpose of building up an automated workflow for the tomographic data processing and analysis. The L-curve based CGTV reconstruction will be applied to two sets of experimental, biological data (imaging data of living weevil and frog embryo) and display high quality reconstructions with limited projection data.

5.1. Parameter optimization

This thesis focuses on the reconstruction problem of insufficient projection data which is seriously ill-posed due to the data insufficiency and measurement errors. A regularization is taken into account to integrate the image property of sparsity for controlling the direction in which the solution is going. That is, a unique solution out of many possible ones is constructed subject to the additional conditions specified by the regularization term. However, the regularization of the inverse problem in (4.16) suffers from a trade-off between the regularized solution and the consistence with the measured data. This trade-off is controlled by the determination of a proper value of the regularization parameter λ.

The regularization is important in solving the inverse problem because the result simply from least square criterion is seriously affected by the data errors and data insufficiency. These errors and the consequent artifacts are reduced by introducing regularization. However, if the regularization is weighted too much, the solution is not consistent with the measured projection data any more as the data fidelity term $\|\mathbf{Ax} - \mathbf{p}\|_2^2$ is is too large, generally leading to an over-smoothed $\mathbf{x}$. On the other hand, if the regularization is weighted too small, the data fidelity will be good but data errors and insufficiency will be more leading to a solution with noise and artifacts. Conceptually, the proper value of the regularization parameter is decided by the applied prior information, such as the sparsity, which differs for different datasets and representing domains. The sparser the object is, the larger the value of λ should be.

Several strategies have been proposed for determining the optimal value of the regularization parameter in the literature such as the discrepancy principle (DP) [60], the generalized cross validation (GCV) [61], the L-curve method [59] and the unbiased predictive risk estimator (UPRE) [115]. The DP strategy is a posterior method which chooses the regularization parameter according to the error level. Namely, the error level of the input measurements must be known. This is not practical in the our experimental tomographic data. GCV is a technique that estimates a regularization parameter directly without requiring an estimate of the error level. It is based on the concept of prediction errors. The basic idea is to take kth observation out of all observed data, and then use the remaining observations to predict the kth observation. If the regularization parameter is a good choice, the kth component of the fitted data should be a good predictor for kth observation on average. However, due to the non-linearity of the TV formulation, the GCV evaluation formula can not be derived explicitly. The L-curve method is mainly based on the plot of the norm of the regularized solution versus the norm of the corresponding data fidelity residual. The method is used to choose a regularization parameter related to the characteristic L-shaped "corner" of the plot. Recently, Lin and Wohlberg [115] extended the method of unbiased predictive risk estimator (UPRE) to the TV regularization problem by considering a quadratic approximation to the TV term. However, they reported the experimental results only on a small problem.

Among the above-mentioned strategies for determining the regularization parameters, the L-curve method has drawn the most attention. It is based on the plot of the regularization term $R(\mathbf{x}_\lambda)$ versus the data fidelity term $\|\mathbf{Ax}_\lambda - \mathbf{p}\|_2^2$ with respect to the regularization parameter λ, forming an L-shaped curve called L-curve as shown in Fig. 5.1a. The corner (vertex point) simply divides the L-curve to two parts: the vertical and horizontal curves. The vertical curve corresponds to relatively small values of the parameter λ, which leads to good data fidelity as the term $\|\mathbf{Ax}_\lambda - \mathbf{p}\|_2^2$ is very small but artifacts might be serious in the solution. The horizontal curve is obtained from larger values of the parameter λ where the regularization becomes overwhelming, may resulting in blocky, over-smoothed solutions. The optimal value for the regularization parameter is generally considered to be located at the corner of the whole curve since at that point both the data fidelity and
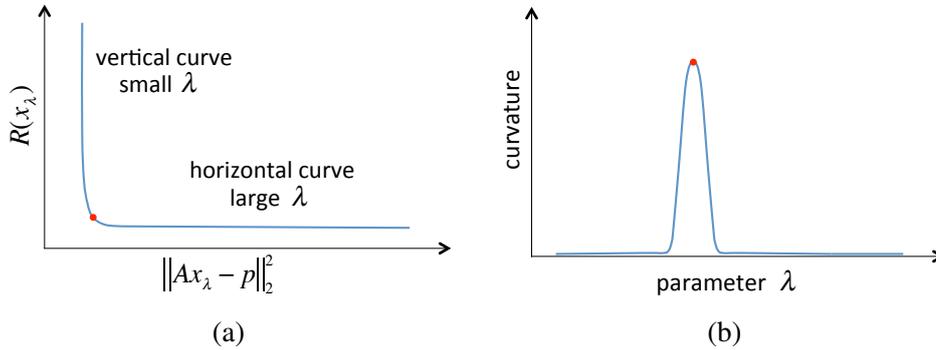
Fig. 5.1. L-curve strategy for the selection of regularization parameter. (a) a schematic diagram for the theoretical L-curve including the vertical and horizontal parts roughly divided by the corner (red vertex point); (b) the function approximately representing the curvature of the L-curve in (a).

regularization terms are relatively small simultaneously and it has the maximal curvature as shown in Fig. 5.1b. The L-curve strategy has been applied to the regularization analysis by computing the singular value decomposition (SVD) of the system matrix $\mathbf{A}$ and the maximal curvature for finding the corner [59]. However, the intensive computation required by the singular value decomposition prevents its application to the large problems such as the tomographic reconstruction. Therefore, a different method is required for finding the corner of the L-curve.

Despite the limitations of the L-curve strategy [116], this idea can be simply employed to the inverse reconstruction problem with TV regularization discussed in this chapter. For a comprehensive analysis of this method in the context of inverse problems see [116].

5.2. Discrete L-curve method to fix TV regularization

As already mentioned, it is advantageous to fix parameter λ in problem (4.16) with a reasonable criterion other than the strict imposition of $\|\mathbf{A}\mathbf{x} - \mathbf{p}\|_2^2$ in the sense of an exact resolution of this constraint by the Lagrangian multiplier method. Rather we would like to maintain a definite balance between the data fidelity and the regularization term.

Different values of λ change the reconstruction quality. This can be seen from Fig. 5.2 which displays the results for the Shepp-Logan phantom of image size 256×256 reconstructed from 60 projections using the CGTV solver for various values of λ. If λ is small (Fig. 5.2a and 5.2b) then reconstruction retains a good data fidelity but streakline artifacts do appear in the reconstructed image. For λ approaching the value 2 (Fig. 5.2c) artifacts

(a) $\lambda = 0.001$ (b) $\lambda = 0.1$ (c) $\lambda = 2$

(d) $\lambda = 16$ (e) $\lambda = 64$
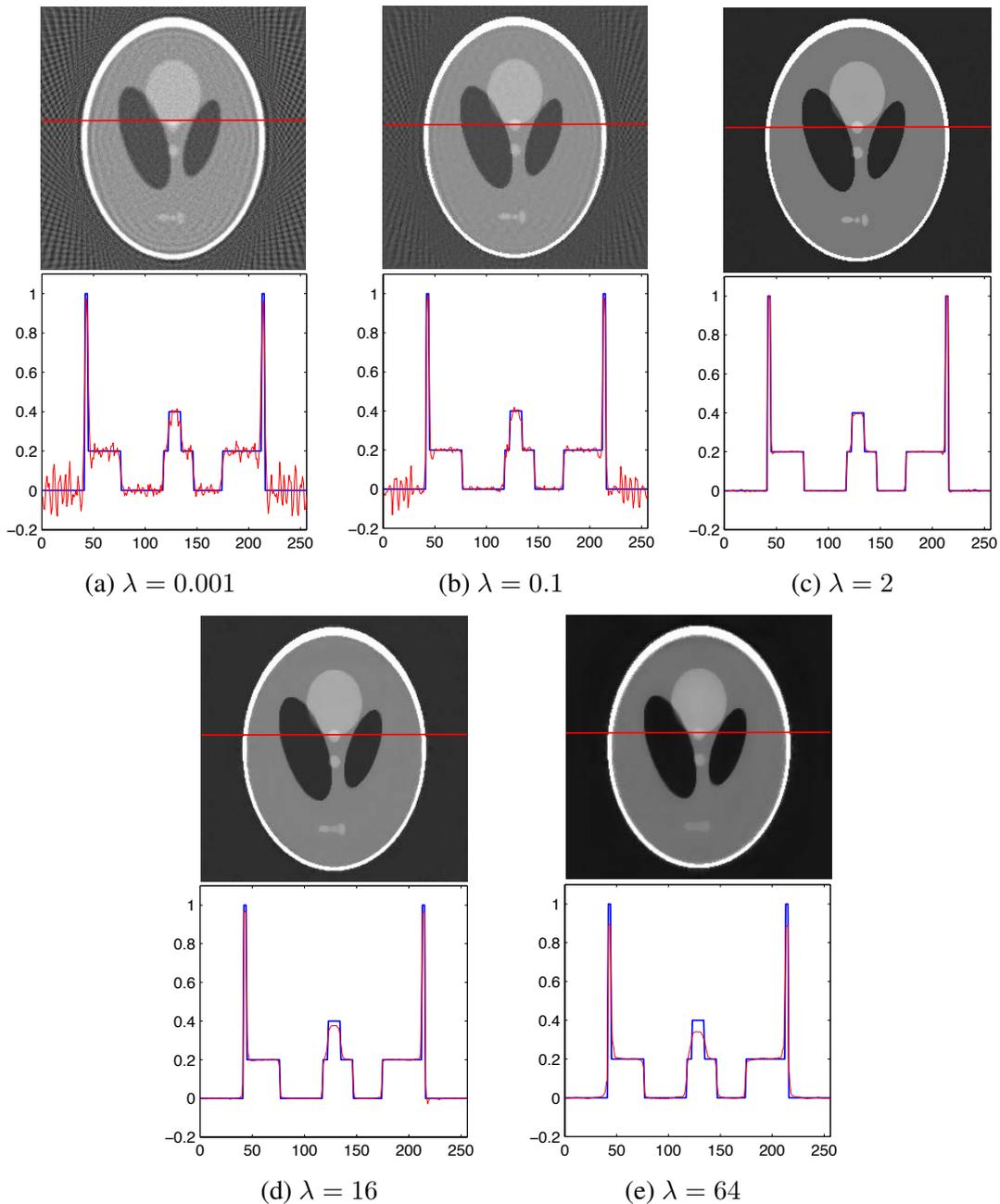
Fig. 5.2. *Upper row:* reconstruction results for Shepp-Logan phantom (256×256) from 60 projections (403 projections required for a high quality FBP reconstruction) using the CGTV solver subject to various values of λ. *Lower row:* Cuts along horizontal through respective images of upper row (marked by red lines). Corresponding cuts through the ground truth are blue lines.
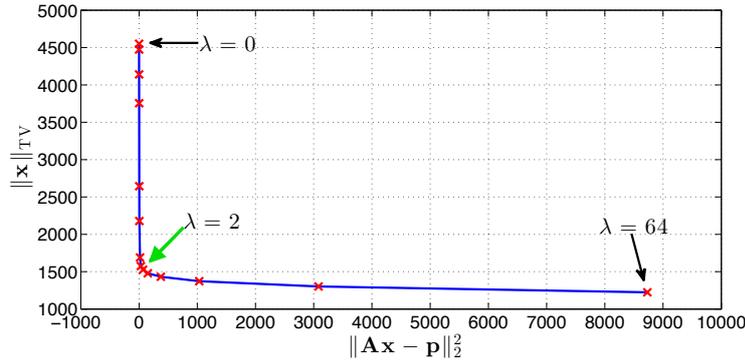
Fig. 5.3. Discrete L-curve for the reconstruction of the Shepp-Logan phantom (256×256) from 60 projections with respect to $\lambda = 0, 0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 1, 2, 4, 8, 16, 32,$ and 64 (for information preserving FBP reconstruction 402 projections would be required). Green arrow indicates the optimal point on the L-curve, corresponding to $\lambda = 2$.

disappear. For larger values of λ (Fig. 5.2d and 5.2e) images becomes oversmoothed, thus losing details around edges. Therefore, it can be visually concluded that a good choice is $\lambda = 2$. An appropriate choice of λ is significant for obtaining a good quality reconstruction.

Let us now investigate whether there is a way to automatically optimize λ. An established method to do this is the use of the L-curve criterion which does not require any prior information. Extensively used by Tikhonov in truncated singular value decomposition methods [117], the L-curve strategy is surely applicable to CGTV. This strategy evaluates the norm of the regularization term versus the data fidelity term in parametric dependence on λ. In our case this corresponds to plotting the parametric curve $\|\mathbf{x}\|_{\mathrm{TV}}$ versus $\|\mathbf{A}\mathbf{x} - \mathbf{p}\|_2^2$. If the regularization term is overweighted, the algorithm reconstructs the object with a low regularization and a high fidelity term. On the other hand, for small values of λ the fidelity term is small while the regularization term is large. As a consequence, parametric fidelity- versus regularization-term dependence is L-shaped, see Fig. 5.3. One may represent this L-curve on a log-log scale. However, the L-shape persists when representing the curve on a linear-linear scale. In our work residuals depend only power-like on λ, and thus a linear-linear representation of the L-curve is used. It is expected that the data fidelity term is small indicating good data consistence and meanwhile the regularization term has also small value indicating the solution being sparse with minimal artifacts. The point in the L-curve satisfying this two expectations is the corner.

Now the question is how to search the location of the L-curve corner. Numerically, it is appropriate to discuss the discrete L-curve method which is based on an interpolation of discrete points on the curve obtained from a finite set λ-values at a fixed number of iterations for the conjugate gradient minimization where stagnation starts to set in. The

L-curve strategy now states that both the regularization and fidelity terms are comparably small. Intuitively, it is clear that the corner of the L-curve meets this requirement. But how can one identify the L-curve corner mathematically speaking? Two criteria can be considered [59]. First, one may seek the point of shortest distance to the origin. The definition of distance could vary for different methods and applications. Alternatively, one may pick the point of maximum curvature. Numerically, a reliable calculation of curvature requires a higher-order-polynomial interpolation. This option is not pursued in this thesis and we resort to the first possibility, minimizing the Euclidean distance $(F(\mathbf{x})^2 + T(\mathbf{x})^2)^{1/2}$.

For the L-curve in Fig. 5.3 reconstructions of the Shepp-Logan phantom in Fig. 5.2 were performed from 60 projections using the CGTV solver subject to 14 values of λ. Breaking the algorithm off after 200 iterations, $F(\mathbf{x})$ and $T(\mathbf{x})$ were calculated and parametrically plotted. From top right to down left points marked by red crosses correspond to the following values of λ: 0, 0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 1, 2, 4, 8, 16, 32, and 64. In the sense of minimum distance to the origin the optimal value is $\lambda = 2$, highlighted by the green arrow. The reconstructed image corresponding to $\lambda = 2$ is displayed in Fig. 5.2c. So, indeed, the L-curve criterion generates the same value of λ as identified above visually, indicating its reliability. For the remainder of this thesis we refer to the CGTV solver subject to the discrete L-curve method as *optimised CGTV* [110].

5.3. Strategy for automized reconstruction

Figure 5.4 displays the workflow of the CGTV reconstruction of a volume. To reconstruct 3D objects from noisy intensity projections, various pre-processing steps need to be performed which include flat- and dark-field corrections to eliminate modulations e.g. introduced by beam- and detector-inhomogeneities as well as hot- and dark-pixel filtering. In the case of phase-contrast imaging phase retrieval is required in addition. Depending on the SNR, the application of various masks to Fourier-transformed (processed) intensity data may be required prior to phase retrieval [118].

Strictly speaking, CGTV reconstruction of a volume out of 2D projection data acquired in a parallel beam would necessitate the separate determination of the optimal value λ^* for each of the 2D slices comprising the reconstruction volume. Obviously, this task can be performed in parallel [40] which will be elaborated in the next chapter. Lacking computing power, it often suffices to determine λ^* in a typical slice and subsequently use it for near-to-optimal reconstruction of all other slices. Once the optimal value of the regularization parameter is obtained based on the L-curve method, it will be used as the input to the CGTV solver for a reliable reconstruction result.
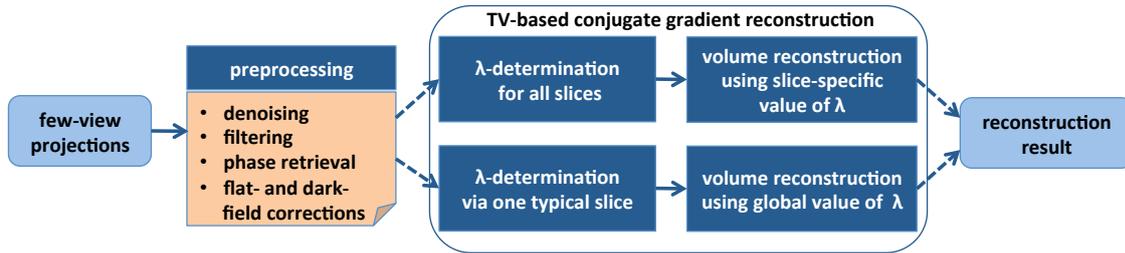
Fig. 5.4. Strategy of volume reconstruction using the optimized CGTV in the case of parallel-beam imaging.

This reconstruction framework is superior in two points. On the one hand, the TV-based conjugate gradient (CGTV) solver does not introduce additional parameters other than the regularization parameter λ, retaining the simplicity of the original problem. On the other hand, the discrete L-curve provides a method for automated parameter determination, independent of expert's experience, producing reliable reconstruction results. However, the parameter determination process is time-consuming. Its incorporation into the framework contributes to a significant increase in the complexity of the reconstruction procedure which will take much longer time for the volume reconstruction even though the parallel reconstruction structure is utilized. This situation is alleviated by the fact that the optimal λ can be calculated once and used for the reconstruction of all parallelized cross-sections.

5.4. Application of optimized CGTV reconstruction

This section will investigate the reconstruction quality of optimized CGTV when applied to a more realistic phantom (photograph of Barbara) and two sets of X-ray *in vivo* data, representing a weevil as well as a stage-17 frog embryo. While reconstructions of the weevil are obtained from a maximum of 400 intensity images, generated by a mix of phase and absorption contrast (no phase retrieval); reconstructions of the frog embryo rely on a maximum of 499 phase maps (projections), retrieved from propagated intensity images. Practically, no absorptive contamination is present in the frog data. The ratio between real decrement and imaginary part of the refractive index at an X-ray energy of ~ 30 keV is $\sim 10^3$ [119]. Thus, judging from the vantage point of image formation, reconstructions of the frog embryo use a lower dose than the reconstructions of the weevil do. The performance of optimized CGTV to that of conventional FBP is also compared.

5.4.1. Image quality assessment (IQA)

Subjective evaluation, which draws upon experience, is often used for CT reconstruction to assess its quality. In general, an analysis of research data should not rely on subjective evaluation. Reproducible results of low prejudice can only be generated by subjecting data assessment to sufficiently general scientific principles (a set of error metrics optimized to given noise types and the effects of missing input information). This is important for meaningful image analysis subsequent to the reconstruction procedure. For example, reconstructed CT data acquired in entomology or developmental biology serve as input to segmentation of tissues and cells, volumetry, cell-mass determination, estimates on the statistics of model parameters, feature extraction, inference of force exertion, flow-field analysis, etc., and each of these procedures propagates the reconstruction error.

3D reconstruction of tomographic data acquired in a parallel beam decomposes into 2D reconstruction of independent slices. Therefore, only a 2D assessment of reconstruction quality is considered here. The simplest and most widely used quality measure is the mean square error (MSE), defined as the average of the squared grey-value differences over all pixels representing reconstructed ($\mathbf{X}$) and reference ($\mathbf{Y}$) slices

$$\text{MSE}(\mathbf{X}, \mathbf{Y}) = \frac{1}{n^2} \sum_{i,j=1}^{n} (\mathbf{X}_{i,j} - \mathbf{Y}_{i,j})^2 \,. \tag{5.1}$$

An assumption behind using MSE as an error measure is that image pixels are statistically independent. Thus MSE exhibits a certain sensitivity to structural changes [120]. A low value of $\text{MSE}(\mathbf{X}, \mathbf{Y})$ indicates good similarity between $\mathbf{X}$ and $\mathbf{Y}$. However, congruence between extended regions is detected more poorly.

Therefore, another image quality index called structural similarity (SSIM) is considered here for complementary understanding of the image [121]. SSIM is based on the fact that images acquired of real objects are usually structured in a hierarchical way, and correlations between grey values in separated pixels occur. The SSIM index is a product of three factors based on a luminance measure $l(\mathbf{X}, \mathbf{Y})$, a contrast comparison $c(\mathbf{X}, \mathbf{Y})$, and association $s(\mathbf{X}, \mathbf{Y})$,

$$\text{SSIM}(\mathbf{X}, \mathbf{Y}) = l(\mathbf{X}, \mathbf{Y})^\alpha \times c(\mathbf{X}, \mathbf{Y})^\beta \times s(\mathbf{X}, \mathbf{Y})^\gamma \,, \tag{5.2}$$

where α, β, and γ are real-positive trade-off parameters which adjust their relative importance. For simplicity, here we set them equal to unity, $\alpha = \beta = \gamma = 1$. The reader is referred to [121] for explicit representations of l, c, and s. SSIM ranges within [0, 1]. Values close to unity (zero) indicate a great (small) structural similarity between the reconstructed and the reference image.
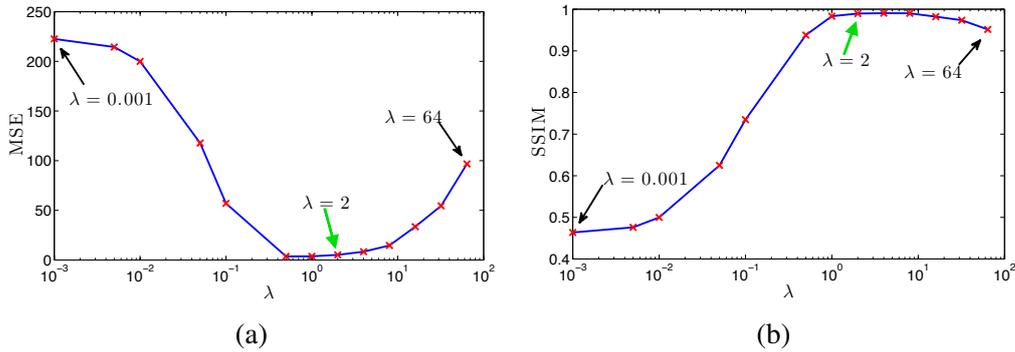
Fig. 5.5. MSE in (a) and SSIM in (b) upon evaluating the CGTV reconstruction of the Shepp-Logan phantom with fourteen different values of λ. Green arrows indicate the optimal value $\lambda = 2$ as determined from the discrete L-curve in Fig. 5.3.

More image quality assessment measures, that are subjective or objective, with or without reference, can be found in the literature [122]. They, however, are not used in this thesis. The two measures MSE and SSIM complementarily work together here to show the image quality in terms of independent pixels and structrual information respectively.

5.4.2. Reconstruction of simulated data

The discrete L-curve of the Shepp-Logan reconstruction in Sec. 5.2 was represented by fourteen values of λ. Figure 5.5 now shows the dependence of MSE and SSIM on the same λ values, demonstrating consistency with the discrete L-curve of Fig. 5.3 in the sense that the optimal value of $\lambda = 2$ is close to the minimum (maximum) of the MSE (SSIM) curve. Note that compared to the L-curve criterion the minimum (maximum) of the MSE (SSIM) curve yields a slight underestimation (overestimation) of this parameter value. This demonstrates consistency of the L-curve method in the sense of a compromise between MSE and SSIM based optimization.

Let us now apply this analysis to a more realistic phantom – the Barbara image (256×256) shown in Fig. 5.6a. In many X-ray CT applications the gradient image is sparse only in an approximate sense. Namely, it is not guaranteed that the object, like the one in Fig. 5.6a, enjoys the property of piece-wise constancy across its entire support as was the case for the Shepp-Logan phantom. Using the same forward model, see Sec. 2.5, 120 projections are sampled which are subsequently taken as input for the CGTV solver subject to the discrete L-curve criterion to optimally reconstruct the image.

The first step is to determine the optimal value of λ. Again, the same fourteen values of λ as used in case of the Shepp-Logan phantom are employed. Figures 5.6b–5.6d display reconstruction results for three different values of λ. Small values of λ, say $\lambda = 0.001$,

(a) Barbara   (b) $\lambda = 0.001$ (c) $\lambda = 4$ (d) $\lambda = 64$
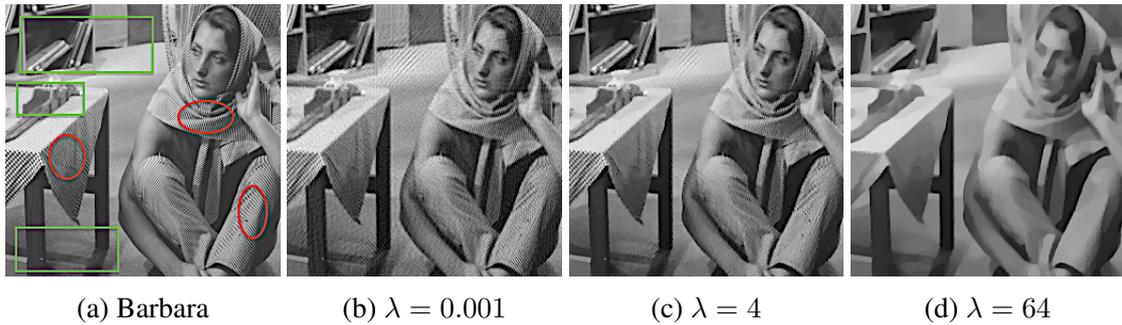
Fig. 5.6. (a) original Barbara (256×256) and reconstruction results from 120 projections using the CGTV solver with (b) $\lambda = 0.001$, (c) $\lambda = 4$, and (d) $\lambda = 64$.
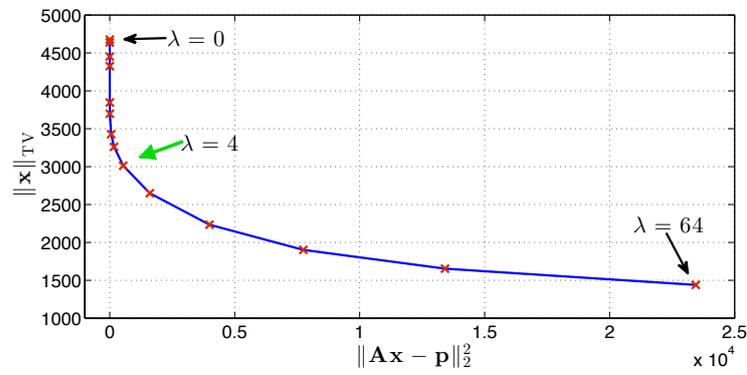

Fig. 5.7. Discrete L-curve for the reconstruction of the image Fig. 5.6a (256×256) from 120 projections (for information preserving FBP reconstruction 402 projections would be required). Values of λ, which are employed, read $\lambda =$0, 0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 1, 2, 4, 8, 16, 32, 64. According to the criterion of shortest distance to the origin, the optimal value is $\lambda = 4$, indicated by the green arrow.

generate noticeable streakline artifacts. In contrast, when λ goes large, details at edges are oversmoothed. The corresponding L-curve is shown in Fig. 5.7. The smallest distance to the origin occurs at $\lambda = 4$, corresponding to the reconstruction in Fig. 5.6c. Theoretically, regions represented by high frequencies (texture) can only be smoothly interpolated from a few projections when using TV regularization due to lack of sparsity in the gradient-space representation. As shown in Fig. 5.6a, areas marked by red ellipses lose texture information in the reconstruction of Fig. 5.6c, but sparse areas, marked by green rectangles, preserve their edges well.

Figure 5.8 depicts reconstructions of the Shepp-Logan phantom and Barbara with a much small number of projections (60 as opposed to 402 in former, 120 as opposed to 402 in latter case) than required for FBP reconstruction without loss of resolution relative to that imposed by pixel size. Note the streaklines in FPB reconstructions subject to these lower number of projections while optimized CGTV is void of such artifacts. Table 5.1,

(a) True image   (b) FBP   (c) CGTV
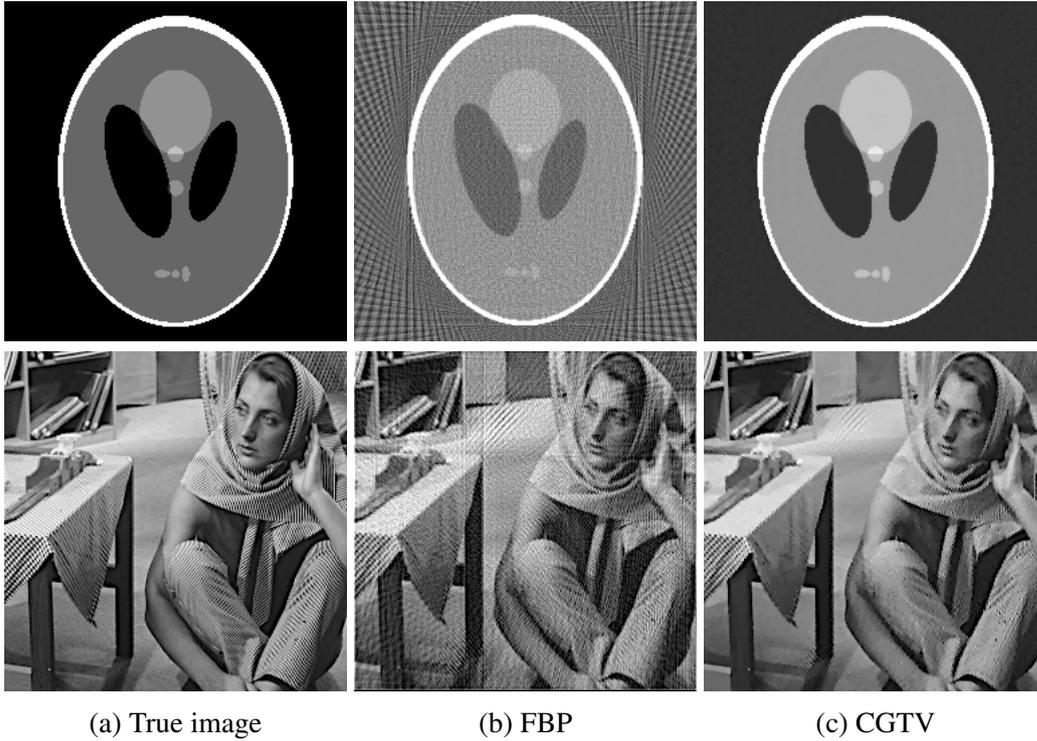
Fig. 5.8. True image and image reconstructions using FBP and CGTV for the Shepp-Logan phantom (first row, 60 projections) and Barbara (second row, 120 projections).

in terms of IQA indexes, shows that optimized CGTV performs much better than FBP for a limited number of projections. Note that due to its piece-wise constancy optimized CGTV reconstruction of the Shepp-Logan phantom is essentially perfect. This is less so for the optimized CGTV reconstruction of Barbara which, however, still performs much better than FBP.

The robustness of the reconstruction results based on the optimized CGTV is further demonstrated in Fig. 5.9 by comparing them to the reconstructions of TVAL3 and TwIST solver [49, 46]. They are implementations of the few-view CT reconstruction problem by using also the total variation based regularization for artifacts reduction. But the parameters in these solvers are determined by the researcher's experience. Many attempts for the parameters are previously performed in order to select the 'best' reconstruction result visually judged by the researcher. The manually determined reconstruction results of TVAL3 and TwIST for the Shepp-Logan phantom and the Barbara image are shown in Fig. 5.9. The CGTV solver with the automated parameter determination process simply provides not worse and even better reconstruction image than TVAL3 and TwIST. This can also be seen from the IQA measures of MSE and SSIM in Table 5.1. Thus the optimized CGTV solver with the assistance of the discrete L-curve method achieves good reconstruction comparable to the subjectively evaluated result in an automated way.

| (a) TVAL3 | (b) TwIST | (c) TVAL3 | (d) TwIST |
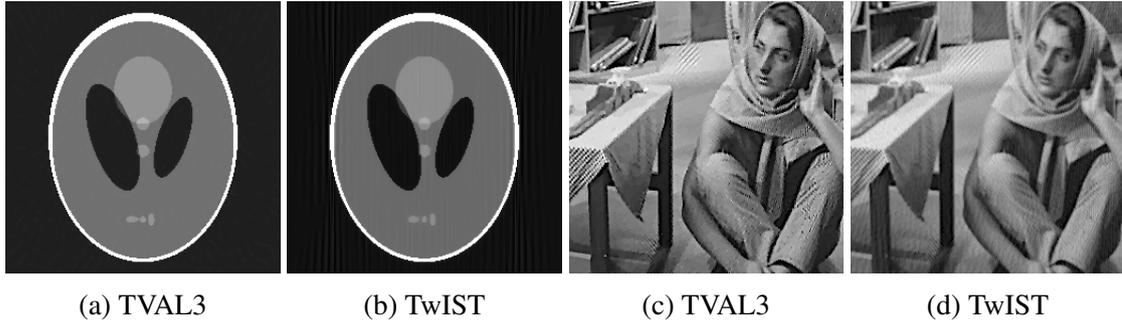
Fig. 5.9. Reconstructions using TVAL3 and TwIST for the Shepp-Logan phantom (60 projections) and the Barbara image (120 projections).

Table 5.1. IQA measures for reconstruction using FBP, TVAL3, TwIST, and optimised CGTV subject to a low number of projections through the Shepp-Logan phantom and the Barbara image.

IQA Measures	Shepp-Logan				Barbara			
	FBP	TVAL3	TwIST	CGTV	FBP	TVAL3	TwIST	CGTV
MSE	843.35	**0.71**	5.79	4.54	867.11	308.37	351.37	**299.05**
SSIM	0.21	**0.99**	0.94	**0.99**	0.40	0.74	0.70	**0.75**

5.4.3. Reconstruction of experimental data

In this section, two different *in vivo* data sets are examined using FBP and optimized CGTV reconstruction, one representing a weevil imaged by a mixture of phase and absorption propagation-based X-ray intensity contrast subject to direct reconstruction (without phase retrieval), the other one embodying a stage-17 frog embryo during neurulation with a stack of exit-phase maps (projection of real decrement of refractive index) serving as input to the reconstruction. Phase retrieval is performed by a variant of the quasiparticle algorithm proposed in [16] and [17] with the high-frequency part of the intensity contrast being cut off at a radial point where noise starts to exceed the signal. Both data sets are subject to statistical noise, in the former case directly present in the intensity data while the stack of phase maps carries the noise contamination in an implicit way. Projections are constrained to the angular range [0°, 180°]. An acquisition of 400 intensity projections was performed per tomogram for the imaging of a weevil. For the frog embryo data, the number of projections was 499. For more detailed information of the imaging of these two datasets, see Appendix A.

In order to perform few-view reconstructions in case of weevil imaging, one fifth of angular, equidistantly-spaced projections (80) are extracted out of 400 projections, while in the frog case a third (167) equidistant projections are taken as input for the reconstruc-
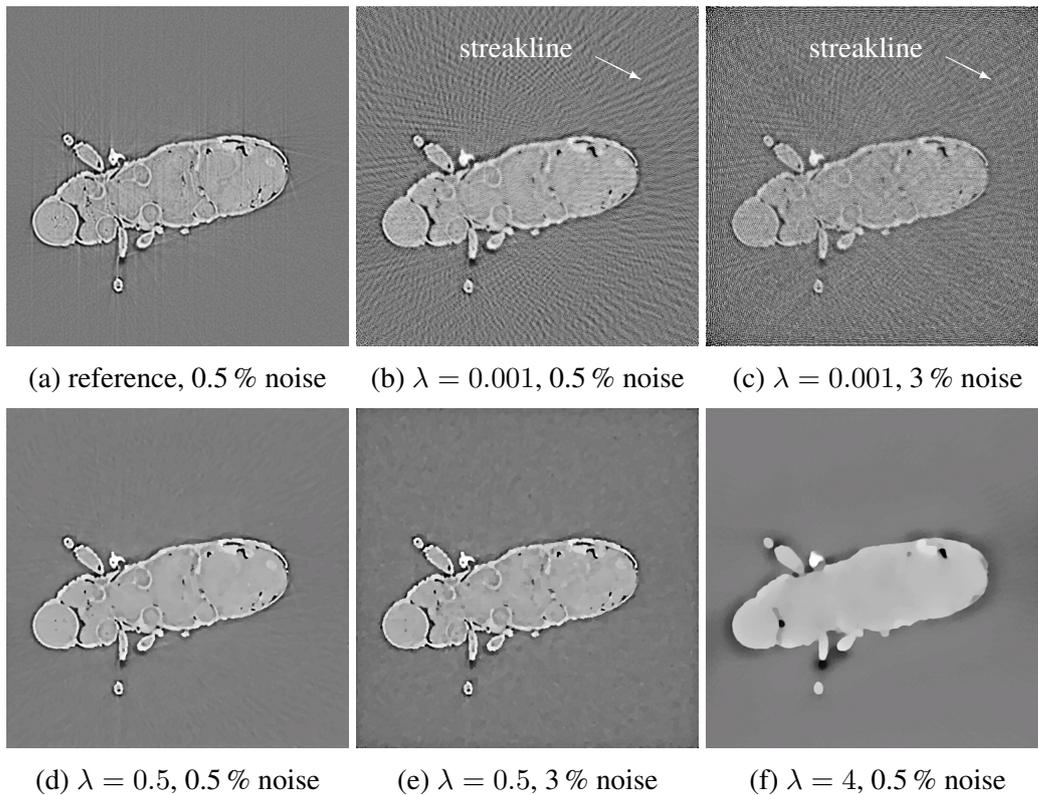
Fig. 5.10. Reference image and λ dependent reconstruction results from intensity data of weevil subject to different levels of Poisson noise and 80 projections: (a) FBP reconstruction using full set of low-noise projections; (b) and (c) CGTV reconstructions for a small value of $\lambda = 0.001$ for 0.5 % and 3 % noise data, respectively; (d) and (e) respective CGTV reconstructions for $\lambda = 0.5$ which is optimal for both, 0.5 % and 3 % noise data; (f) CGTV reconstruction using $\lambda = 4$ for 0.5 % noise data.

tion. This choice of data thinning is motivated by Figs. 5.16 and 5.17 to be discussed below. Sparsity holds fairly well for the weevil data (chitin skeleton, large cavities represent piece-wise constant regions), thus few-view reconstruction should lead to good results. In contrast, the frog embryo exhibits, at a comparable imaging resolution, finer structures (cells of variable size, cell nuclei, yolk platelets, boundaries between densely and loosely packed cells) leading to appreciable variations within regions. In principle, more projections thus are necessary for faithful reconstruction.

Thanks to parallel-beam illumination, we focus on the reconstruction of one 2D slice only. Recall that the reconstruction of the 3D volume is accomplished by simple vertical stacking of such 2D slices, allowing parallel computing for fast reconstruction even though this produces an asymmetric treatment of sparsity in x, y versus z gradients. Reference slices are FBP-reconstructed using the full number of projections.
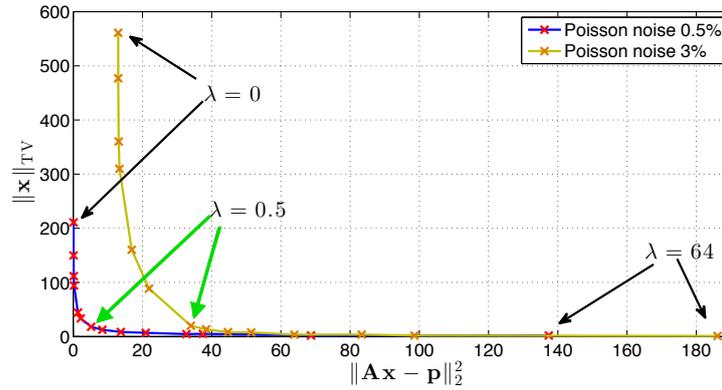
Fig. 5.11. Discrete L-curves for CGTV reconstruction from 80 projections of the weevil data with 0.5 % (blue curve) and 3 % (lime green curve) Poisson noise using $\lambda = 0$, 0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 1, 2, 4, 8, 16, 32, 64. Both curves exhibit minimum distance to the origin at the same value $\lambda = 0.5$.

Figures. 5.10a, 5.10b, 5.10d, 5.10f display weevil-reconstruction results in one and the same slice using CGTV subject to various values of λ. The level of Poisson noise in the intensity projections here is about 0.5 % [12]. Such a low noise-level is expected to propagate to the reconstructed object, maintaining the same order of magnitude. Statistical noise thus is negligible for weevil reconstruction based on the original data. To check exemplarily how a substantial increase of Poisson noise affects the CGTV reconstruction a 3 % noise level (6 times as the original) is artificially imposed on the average intensity of weevil data. The corresponding discrete L-curves are shown in Fig. 5.11 for both cases of 0.5 % and 3 % Poisson noise. Minimum distance to the origin occurs for both at the same value $\lambda = 0.5$ (green arrows in Fig. 5.11). Notice that for $\lambda = 0.5$ the values of $\|\mathbf{x}\|_{TV}$ are comparable while $\|\mathbf{Ax} - \mathbf{p}\|_2^2$ is considerably larger for the case of 3 % Poisson noise. This indicates that optimal CGTV reconstruction universally annihilates small-scale fluctuations albeit subject to a mild alteration of the large-scale structure, comparing Fig. 5.10d with Fig. 5.10e. However, a shift of the optimal value of λ in the right direction is expected if further higher level of Poisson noise is added. Obviously, if λ is small there are streakline artifacts (Fig. 5.10b) and pronounced decrease in quality for the high noise data (Fig. 5.10c). Streakline artifacts are absent in Figs. 5.10d and 5.10e (optimal reconstruction) and 5.10f, the latter, however, exhibiting poor resolution because of higher Poisson noise.

Figures 5.12b–5.12d display according CGTV reconstructions of the stage-17 frog embryo. The level of Poisson noise in intensity projections is comparable to the original weevil data, roughly of the order of 0.3 %, which is low. Note, however, that towards the bottom part of the slice there is motion induced blur (systematic error) due to an onset of developmental dynamics during the tomographic scan. The L-curve is shown in Fig. 5.13 with the minimum distance to the origin occurring for $\lambda = 0.1$ (green arrow in Fig. 5.13).

(a) reference   (b) $\lambda = 0.001$ (c) $\lambda = 0.1$ (d) $\lambda = 4$
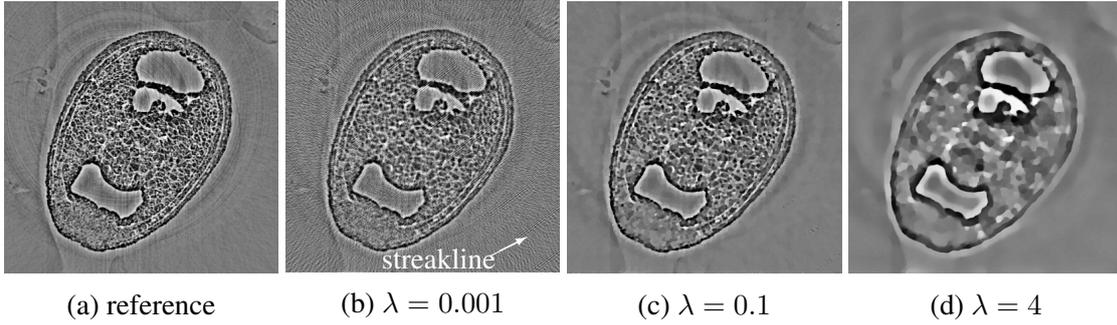
Fig. 5.12. Reference image and λ dependent reconstruction results from phase maps retrieved from intensity projections of stage-17 frog embryo: (a) FBP reconstruction using full set of projections; (b)–(d) CGTV reconstructions using 167 projections for different values of λ.
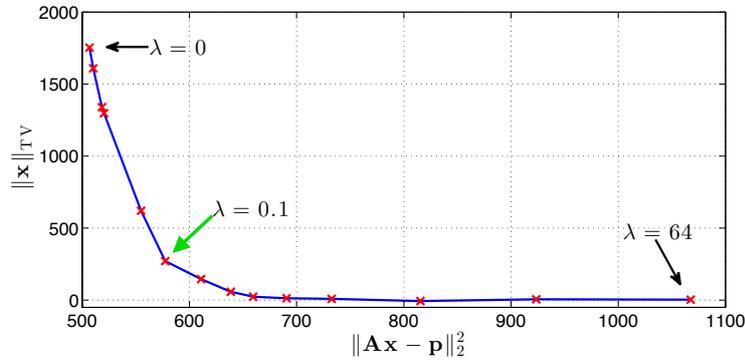

Fig. 5.13. Discrete L-curve for CGTV reconstruction from 167 phase maps of stage-17 frog embryo using $\lambda = 0, 0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 1, 2, 4, 8, 16, 32, 64$.

The associated reconstruction is depicted in Fig. 5.12c. Again, streakline artifacts occur in Fig. 5.12b, which are absent in Figs. 5.12c (optimal reconstruction) and 5.12d.

Interestingly, by visual observation the presence of a lightly weighted TV minimization constraint for few-view CGTV reconstruction appears to mimic FBP reconstruction as far as the occurrence of streakline artifacts is concerned. This is shown in Fig. 5.14. Thus, by a small value of λ the degeneracy in finding the (global) minimum of $\|\mathbf{Ax} - \mathbf{p}\|_2^2$ (underdetermined linear system of equations) is lifted in the sense of unique FBP reconstruction.

To point out the loss of resolution even for optimized CGTV Fig. 5.15b presents a line cut through the frog-embryo slice shown in Fig. 5.15a with reconstructions performed by FBP subject to the full number of projections ($P = 499$, reference, red curve) and optimized CGTV using only $P = 167$ (blue curve). Clearly, the blue line misses high-frequency components (noisy structure in flat regions) which are present in the red line.

85

(a) FBP   (b) CGTV ($\lambda = 0.001$)   (c) FBP   (d) CGTV ($\lambda = 0.001$)
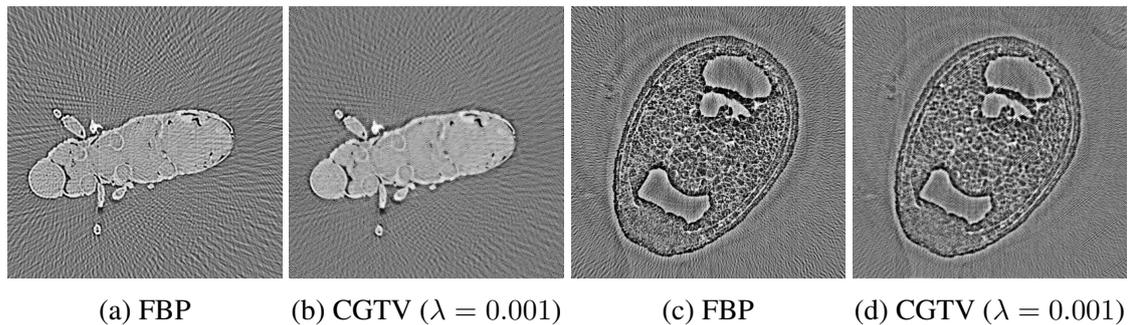
Fig. 5.14. Comparison of FBP and small-λ CGTV reconstruction of weevil and frog-embryo slices ($\lambda = 0.001$ in both cases) using a small number of projections, $P = 80$ and $P = 167$ (as opposed to maximum numbers $P = 400$ and $P = 499$), respectively. Note that streakline artifacts appear for both FBP and small-λ CGTV reconstruction.

There is high-frequency information in the optimized CGTV reconstruction with $P = 499$ (green curve), occasionally albeit not always representing that of the red line. It can be summarized that the low-dose (few-view) reconstruction using optimized CGTV is free of streakline artifacts which necessarily are introduced by FBP. Optimized CGTV thus operates cleanly with an acceptable loss of resolution towards long *in vivo* time-lapse series in the following sense: it depicts, say, cellular shapes by precise reconstruction of cell boundaries, it excludes the appearance of FBP few-view reconstruction artifacts, but it does not resolve small intracellular details like yolk platelets and nuclei. Due to lack of information low dose (few view) reconstruction theoretically has difficulties in recovering high frequency components which destroy the sparsity of the reconstruction image. The sparsity assumption behind CGTV tends to not only remove the streakline artifacts but also cancel details in the reconstruction, however, with preserving important edges. The sparsity of the reconstruction is helpful in the segmentation of structures in different scale. For this purpose the value of λ can be manually adjusted. For dose-intense (many-view) reconstruction of sub-cellular resolution FBP is still preferred due to its computational efficiency, however.

Finally by using the two image quality assessment (IQA) metrics MSE and SSIM as defined in Sec. 5.4.1, a question can be asked that how the reconstruction quality depends on the number of projections P. We take the full-projection FBP reconstruction as a reference. As Fig. 5.16 (weevil) and Fig. 5.17 (stage-17 frog embryo) indicate both IQA measures saliently saturate at $P \sim 80$ and $P \sim 167$ (green arrows), respectively. This justifies our use of these P values in the analysis above.
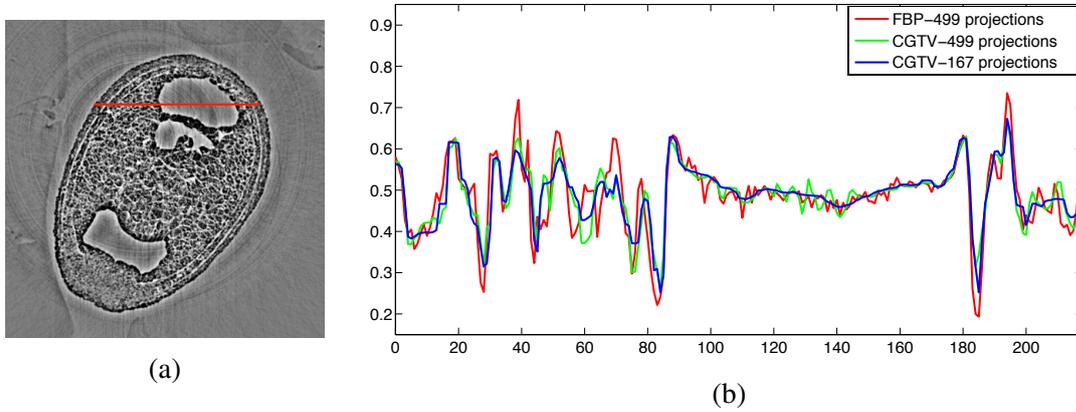
Fig. 5.15. (a) Definition of line through reconstructed slice of stage-17 frog embryo, (b) reconstruction results (normalized to [0, 1]) along this line for FBP subject to the full number of projections ($P = 499$, reference, red curve) and optimized CGTV using $P = 167$ (blue curve) and $P = 499$ (green curve).
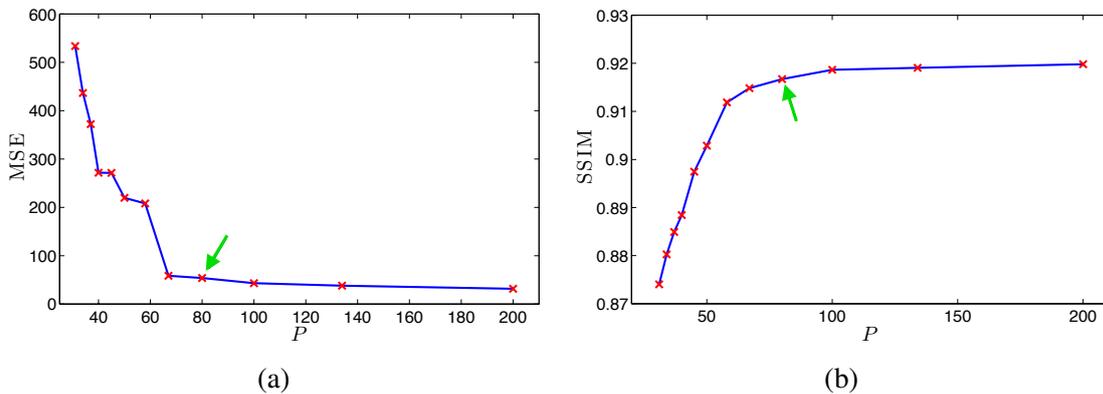

Fig. 5.16. Error of optimized CGTV with respect to full-view FBP reconstruction ($P = 400$) of weevil slice using error metrics MSE and SSIM, defined in Sec. 5.4.1, versus the number of projections P. Green arrows mark approximate onset of saturation.

87
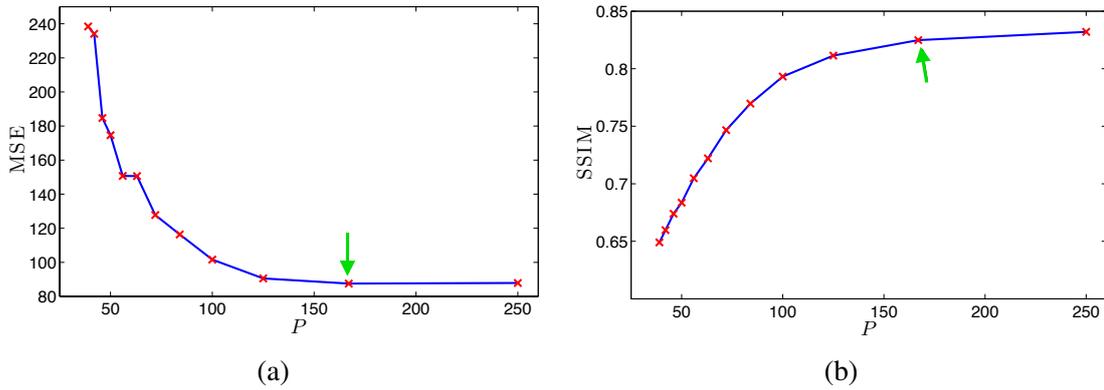
Fig. 5.17. Error of optimized CGTV with respect to full-view FBP reconstruction ($P = 499$) of stage-17 frog-embryo slice using error metrics MSE and SSIM, defined in Sec. 5.4.1, versus the number of projections P. Green arrows mark approximate onset of saturation.

5.4.4. Convergence analysis and computational performance

To demonstrate relatively fast convergence of the regularization of ART by TV-minimization using the conjugate gradient solver, the behavior of the objective function for the data set of weevil is depicted in Fig. 5.18, corresponding to (4.16), in dependence of the iteration number k for the weevil data set. Five curves corresponding to different values of λ are shown, in which red denotes $\lambda^* = 0.5$. The blue and green curves are almost overlapped, thus it is hard to read the blue one. For λ values in the vicinity of λ^* no appreciable decent occurs for $k > 30$.

Regarding computational performance, forward and backward operations (such as $\mathbf{Ax}$, $\mathbf{A}^{\mathrm{T}}\mathbf{x}$) are required and take a considerable time in each iteration. Typically, the computational time required for a CGTV volume reconstruction with a given value of λ takes about 25 min employing the parallel reconstruction strategy and hardware configuration in next chapter.

5.5. Summary

In this chapter, a particular iterative reconstruction technique (optimized CGTV) is proposed and evaluated. It minimizes the regularized total variation (TV) in a Lagrangian multiplier formulation using the conjugate gradient solver. The optimal value of the regularization parameter is automatically obtained by applying an independent L-curve strategy. By 'independent' it means that the regularization parameter is not empirically
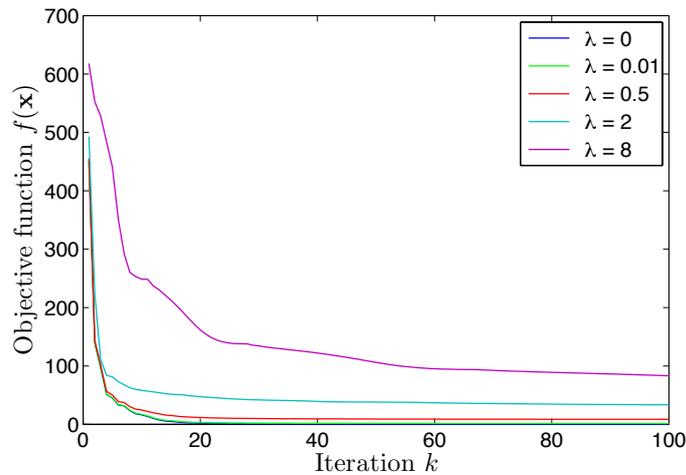
Fig. 5.18. Convergence analysis of CGTV reconstruction of weevil data for five different values of λ. The optimal value is $\lambda^* = 0.5$ denoted in red, and in its vicinity minimization saturates for a number of iterations $k > 30$.

determined by resolving the Lagrangian multiplier formulation. Rather, a physically reasonable flexibility is introduced doing justice to the principle of least prejudice on the reconstruction result. This work is a step of key importance in building an automated workflow for tomographic data processing and analysis.

By applying the optimized CGTV to reconstruct phantom and *in vivo* data which are acquired by propagation based X-ray imaging with and without phase retrieval, it can be concluded that the few-view reconstruction represents a promising venue for fast tomography and low-dose tomographic imaging. Namely, it maintains resolution at an acceptable level, continues to realistically reproduce edges which bound sufficiently large structures (cells, tissue confines), quickly saturates the image quality as the number of projections increases, and is void of streakline artifacts as generated by filtered back-projection (FBP) reconstruction. This is important, e.g., for automated image analysis such as structure segmentation, often resorting to a prior set grey-value thresholds. That is, moderately sacrificing the spatial resolution is not a strong limitation if the object in biological research is to understand coarse phenotypic dynamics: tissue separation, cell division and migration, and cavity formation. All these only require a clear segmentation of the associated boundaries. Thus, the optimized CGTV should be interesting for 4D *in vivo* imaging, aspiring to the acquisition of large amount of time-lapse volume data [11, 12, 14, 15].

The promising reconstruction results of experimental data demonstrate that the L-curve based CGTV reconstruction method in this chapter is practically applicable to the 4D *in vivo* tomographic imaging in study of biological objects. Moreover, it can be applied to other tomographic applications and task-specific reconstruction problems that resort to some sort of sparsity.

6. Iterative reconstruction at the LSDF

Even though the computed tomography reconstruction with limited number of projections has greatly reduced the amount of data involved in computing, iterative reconstruction is still very computing- and memory-intensive in practical applications. In this chapter, the integration of the iterative reconstruction method into a complete reconstruction framework will be discussed. For this an automated workflow connecting the tomographic experiment station to the Large Scale Data Facility (LSDF) has been built up in order to enhance the data storage and analysis efficiency. The high computing requirements of iterative 3D tomography reconstruction will be alleviated by the parallel reconstruction framework at the computer cluster connected to the LSDF. The iterative reconstruction method and the parallel reconstruction framework, after integrated into the workflow, will not only produce promising reconstruction results but also greatly improve the data processing efficiency in various tomographic applications.

6.1. Computing requirements of iterative reconstruction

The usage of iterative reconstruction method was motivated by the requirements of fast tomography imaging and radiation dose reduction. Iterative reconstruction is able to generate high quality reconstruction images even in the case of limited number of projections, as it can easily integrate prior information of the object into the reconstruction procedure by using regularization. Iterative reconstruction, however, is known to be computing-intensive, thus rarely used in practical applications which demand real-time reconstruction performance. In general the purpose of computed tomography reconstruction is to retrieval the structures inside the original object in three dimensions. Two factors can contribute to the intensive computing property of the iterative 3D volume reconstruction.

One is the large amount of projection data in computed tomography. A raw data set for a 3D object consists of a series of 2D projections in different angles. Fortunately, the amount of projection data required for a promising reconstruction can be reduced by the application of iterative reconstruction method. Let us take the tomographic beamline

Topo-Tomo [13] at the synchrotron light source ANKA [33] as an example. One typical data set intensively acquired from a 3D object is 3.3 Gigabytes with around a thousand of projections. Due to the application of iterative reconstruction, the number of projections can be reduced to ~100 which will be only at Megabyte scale. However, if a fourth dimension, time, is taken into account for observing a living object, the amount of raw data will be dramatically increased. It is a challenging task to process the data acquired from the high-speed and high-throughput imaging beamline.

The other factor that makes the iterative reconstruction computing intensive is the feature of the reconstruction algorithm itself. The iterative reconstruction consists of two major computing intensive steps: forward and backward projection. Both steps have to be performed repeatedly in each iteration. Tens of iterations are necessary until a reconstruction result with high image quality is achieved. Furthermore, the iterative reconstruction is also memory-intensive due to the large system matrix even though it is sparse. For a typical reconstruction of a 2D image 1024×1024, the system matrix takes up $\sim3GB$ of memory. The large system matrix is used often in the steps of forward and backward projection and its computation is also time-consuming. It is preferred to store the system matrix in memory instead of computing the matrix whenever it is used. The size of system matrix goes up to Petabyte scale if 3D reconstruction is considered.

The large amount of raw data and the iterative volume reconstruction typically lead to intensive computation in various CT applications. To overcome the limitations and to build real time solutions, it is necessary to use hardware architectures with massive parallel computation capabilities. The use of graphic processing units (GPU) [123, 124] is an appropriate and efficient solution. Due to its highly parallel structure, it is effective for algorithms where processing large blocks of data is processed in parallel. In CT reconstruction GPUs are often used to conduct the computing-intensive steps of forward and backward projections. However, it is difficult to deal with the large memory required by the system matrix. The system matrix has to be calculated whenever it is called. Moreover, in the field of image processing it is convenient to implement algorithms using MATLAB other than high-level programming languages [125] enabled for GPUs. Furthermore, scientific researchers always have the requirements to store and organize the large experimental datasets and make them accessible for a long period of time. In this sense, GPUs are not a good option.

In the implementation of this thesis the author does not resort to any GPU architecture but the Large Scale Data Facility (LSDF) located at the Karlsruhe Institute of Technology (KIT). Connecting the tomographic experiment station to the LSDF will greatly enhance both data storage and processing efficiency. Due to the dramatically advancements in computing technology, general purpose computers have shown the potential to meet the requirements of CT time constraints [126, 22]. Thanks to the parallel beam used for the tomographic data involved in this thesis, the 3D reconstruction can be divided into a series of 2D slice reconstructions, which can be easily parallelized in the computer cluster.
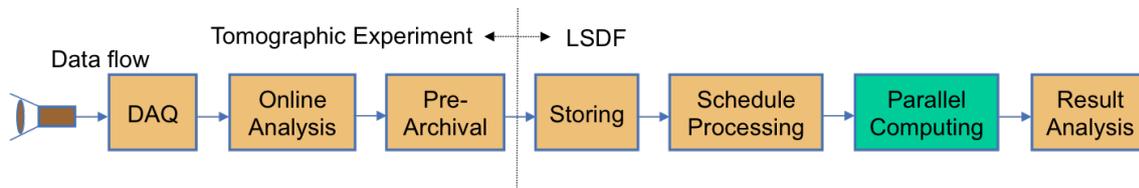
Fig. 6.1. An automated data management and processing workflow designed for tomo-graphic experiments, showing the dataflow from data acquisition to result analysis. It connects the experimental station to the LSDF for utilizing advanced data management services.

Moreover, the computer cluster is able to provide large memory for the system matrix of 2D slice reconstruction and also the MATLAB implementing environment. The parallel computing system with computer cluster is of highly scalability and easily expandable. In the literature hybrid architectures [105, 24] of GPU and computer cluster was also applied in CT reconstruction, but this thesis will only focus on the parallel reconstruction framework using the LSDF computer cluster and its acceleration performance with a 3D reconstruction job.

6.2. LSDF workflow

The LSDF is designed to cope with the increasing requirements of data intensive scien-tific experiments. It is a distributed storage facility at Petabyte scale to support storage, management and access of large volumes of scientific data. Moreover, a computer cluster [127, 23] is connected for data intensive application with a high speed dedicated network infrastructure. For data intensive analysis, the computing performance is typically deter-mined by the setup of LSDF in aspects of hardware and software. More information about the hardware and software setup is described in Appendix B.

Given the provided setup in terms of hardware and software at the LSDF mentioned above, the connection of the tomographic experiment station and LSDF will greatly en-hance the storage, processing efficiency, and result analysis for large volume of datasets acquired from the tomographic experiment. To reduce the complexity of the whole pro-cedure of tomographic data storing, reconstruction and analysis, an automated workflow is designed in Fig. 6.1 showing the dataflow from data acquisition to result analysis.

At the tomographic experiment side, projections firstly are recorded by the camera. Then the data flows through the data acquisition system "DAQ", storing the raw data in a com-puter, and "online analysis", validating the raw data by using a fast reconstruction tech-nique ensuring that the data are congruent with expectation. In the "pre-archival" step, the

experimental and user meta data are integrated with the raw data. Until now, the data are prepared for further storage and processing.

At the LSDF side, the raw dataset is firstly stored into the LSDF data storage resources. At the same time the related meta data, providing such information as the dataset location, user meta data, experimental configuration, processing and monitoring meta data in LSDF, etc., are written into a database. Once the storing process is finished, the monitoring status is written as "ready for processing", which is regularly queried by an execution framework for data intensive applications called LAMBDA [128]. Any new dataset prepared for processing can be detected automatically, which will activate creating a processing procedure in the following step of "schedule processing". This is performed by LAMBDA according to the processing route predefined in the meta data by the user in the step of "pre-archival". The processing procedure is converted to a script executable in the software environment of computer cluster, and then submitted to every node of the computer cluster where the "parallel computing" for the dataset processing is carried out. Finally, the "result analysis" will evaluate the quality of results, store them back into the LSDF storage. Possible failures or errors reported in this step can be handled by users or developers according to their types. The final reconstructed volume, as well as the corresponding raw dataset, are accessible for users in the LSDF via the ANKA repository [129].

This workflow provides an automated strategy for reconstructing the tomographic data by utilizing the capability of LSDF for storage and intensive computing. In the following sections we will only focus on the highlighted step of "parallel computing" in Fig. 6.1 to elaborate the implementation of 3D tomographic reconstruction and the performance achieved by the parallel computing strategy.

6.3. Parallel reconstruction framework

Before we go to deep discussion about the parallel reconstruction, we need to declare that the parallel reconstruction is based on the geometric property of parallel X-ray beams. Namely, a 3D volume reconstruction can be composed of numerous 2D slice reconstructions, each of which is considered to be independent of the others. It is straightforward to perform these slice reconstructions independently in different nodes of a computer cluster. Thus, by "parallel reconstruction" here, we mean the 3D reconstruction parallelized at the data level, which is unrelated to the algorithm employed for the slice reconstruction. Thus the performance achieved by parallel computing should be independent of the applied reconstruction algorithm. Parallelization at algorithm level can further accelerate the reconstruction process. But the algorithm parallelization is not the topic in this chapter. In this sense, we resort to a solver called TVAL3 [51] instead of CGTV in solving the regularized iterative reconstruction problem in (3.29), since the TVAL3 solver was firstly

used for testing the parallel reconstruction performance before CGTV was implemented. Considering that the CGTV solver converges more slowly than the TVAL3, it will take more time for a complete 3D reconstruction job. However, conclusion will be similar for the acceleration performance of the parallel reconstruction framework.

The CGTV solver applies the total variation (TV) as the regularization in the sense of compressive sampling theory, thus being able to reconstruct a promising image from incomplete projection data as discussed in Chapter 3. However, note that the TVAL3 solver, using the augmented Lagrangian method and alternating direction algorithm, introduces more parameters than the regularized form in (3.27) when the constrained minimization problem in (3.29) are mapped into unconstrained one. There total variation is generalized in terms of a variety of constraints (various generalized gradient transforms), formulated in terms of new variables ω_i in addition to $\mathbf{x}$, to give rise to a new minimization problem in ω_i and $\mathbf{x}$. The formal substitution of these additional constraints into the minimization problem gives back the higher dimensional equivalent to the old problem in (3.29). The important insight is to treat the additional constraints in the sense of assigning augmented Lagrangian parameters $\nu_{i,1}$, $\beta_{i,1}$. The augmented Lagrangian method and alternating direction algorithm appear to exhibit superior convergence properties to the solution of the constrained equations. However the determination of these additional parameters are difficult and impractical to apply the L-curve method in Chapter 4. One may just set these parameters to values which visually lead to somewhat expected results (human judgment). This should be avoided for the designed automated workflow as discussed in Chapter 4. But for the investigation of the performance of parallel reconstruction framework, it can be ignored in this chapter considering that the conclusion about the performance are theoretically independent of the reconstruction algorithms.

The TVAL3 solver is an iterative reconstruction approach that calculates the original object image through a loop of updating steps. The 3D reconstruction, if executed by sequentially performing the 2D slice reconstructions, can take tens or even hundreds of hours. In this case the computer cluster connected to the LSDF will play an important role in accelerating the reconstruction procedure, providing the results to users in near real time. In the following sections, we will use a representative dataset to see how the solver works for limited projection data and the performance of a parallel reconstruction framework using the computer cluster connected to LSDF.

6.3.1. Data parallel computing

The data set used in this implementation is the same as the X-ray tomography data described in 4.4.1. The detailed information is given in Table. 4.1. The data set consists of 1024 slices, each reconstructed by the same algorithm TVAL3 with predefined parameters. This reconstruction job is time-consuming if performed sequentially at a general workstation. It is straightforward to distribute this job to a computer cluster in a parallel
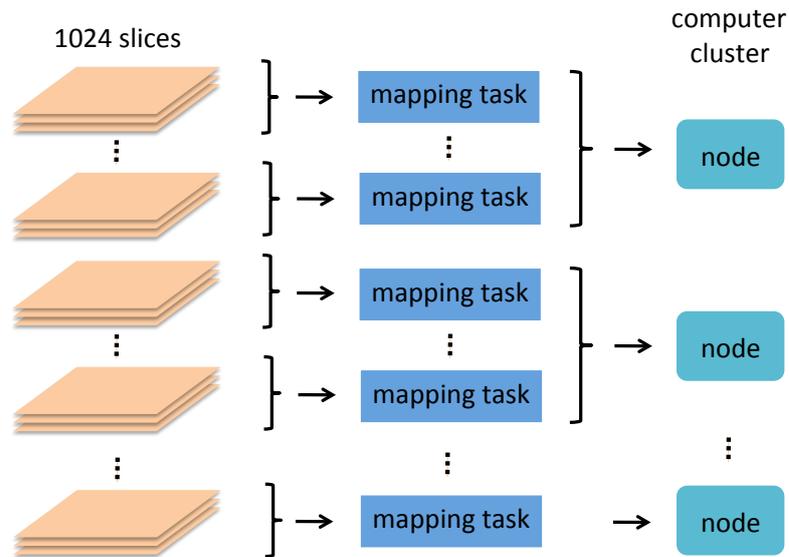
Fig. 6.2. Data parallel computing for 3D computed tomography reconstruction using computer cluster.

way that numerous nodes simultaneously work on different slice reconstructions. The output of these nodes will be stacked to a whole volume reconstruction. In this implementation, 37 nodes in the cluster are available, each of which is capable to process up to six mapping tasks and two reducing tasks in parallel, which is a compromised configuration in Hadoop system between task size and the number of cores required. Thus, as shown in Fig. 6.2 the whole job, 1024 slice reconstructions, can be divided into a number of mapping tasks, which are distributed to the nodes of computer cluster. Note that inside each mapping task multiple slice reconstructions can be performed sequentially, while the mapping tasks are parallel executed in the nodes. Thus the size of the mapping tasks is of key importance for the reconstruction performance.

We expect the size of the mapping tasks as small as possible to avoid sequential computing which takes long time. However, due to the fact that the maximum capability of mapping tasks in the cluster is 222 (37 nodes × 6 mapping tasks per node), each mapping task averagely contains ∼5 slice reconstructions performed sequentially. More mapping tasks (smaller size) will make some of them suspended in a waiting list, which, on the contrary, slows down the speed of reconstruction. If we go to the other direction, setting the size of mapping tasks larger, the number of these mapping tasks will be less than the number of nodes (37). This is not sensible since the computing resources are not fully utilized. The limit of the number of mapping task is equal to one. Namely, all slice reconstructions are grouped in a single mapping task, so that all slice reconstructions are performed sequentially. In this case, the computing performance is comparable to the sequential computing in a general workstation.

Nevertheless, in order to investigate the parallel computing performance of the Hadoop cluster in performing the 3D CT reconstruction, we increase the number of mapping tasks (decrease the size of mapping tasks) gradually from 1 to 222. This can be divided to two stages:

1) *One mapping task per node* stage, in which the number of mapping tasks is smaller than 37 (number of nodes), thus no more than one mapping task is included in each node. In this stage, the number of mapping task is equal to that of nodes actually used for reconstruction. Each additional mapping task will be distributed to the free nodes, thus we expect a fast improvement in computing speed as the increasing of the mapping tasks. The time in second required by the whole reconstruction job can be directly recorded to demonstrate the parallel computing performance in this stage.

2) *Multiple mapping task per node* stage, in which the mapping tasks are more than the number of nodes, making more than one mapping task running in each node. Considering that the nodes in the computer cluster feature dual quad-core processors, the computing power of each node can be better utilized in this stage by the multiple mapping tasks executed in parallel. Compared to the stage of "one mapping task per node", we can expect further improvement in the whole computing performance even though all nodes are already engaged in parallel computing. Since at most six mapping tasks can run in one node simultaneously, we will measure the time in seconds required by the whole reconstruction job with respect to the number of mapping tasks per node. It is clear that the total number of mapping tasks is equal to that per node times 37 (number of nodes).

The time in seconds taken by the whole 3D reconstruction job is measured directly, from which another essential factor, speedup factor, is commonly used to evaluate the performance of parallel computing. The speedup factor is defined as the ratio of time required by sequential implementation to that by parallel computing like

$$\text{speedup}(n) = \frac{\text{time\_sequential}}{\text{time\_parallel}(n)} \, , \tag{6.1}$$

where n is the number of parallelized mapping task. The speedup factor is a convenient indicator of computing speed and efficiency increase in assessing the parallel computing performance.

6.4. Reconstruction and computing performance

Now turn to the computing performance for the 3D reconstruction using the computer cluster. As mentioned above, it is necessary to discuss the two stages separately. For both stages, the reconstruction time in seconds for 3D volume data are recorded and the

Table 6.1. Computing performance in the stage of one mapping task per node

number of nodes n	time (s)	speedup factor
1	39727	1
6	6681	5.94
12	3415	11.63
18	2326	17.07
24	1747	22.73
30	1400	28.37
36	1180	33.66

speedup factors are then calculated. The reconstruction time is measured ten times for every case and the average value is taken. If only one node is used, the 3D reconstruction is a sequential process taking 39727s (~ 11 hours) for all slices. This means that each slice reconstruction averagely needs 38.8s. In order to evaluate the measured reconstruction time and speedup factor, theoretical values for reconstruction time and speedup factor are also involved here. They are calculated theoretically from the time needed by sequential reconstruction and the number of parallelized mapping tasks, representing ideal situations. The theoretical time is equal to the sequential time divided by n. Namely, the parallel computing structure with n mapping tasks is supposed to accelerate the reconstruction process by a factor n. The theoretical speedup factor is equal to the number of mapping tasks.

Table 6.1 gives the the recorded time and speedup factors in the stage of one mapping task per node. As the increase of the number of nodes used, the time decreases rapidly from 37727s to 1180s and speedup factor goes up to 33.66. Fig 6.3 demonstrates the comparison of the measured performance (reconstruction time and speedup factor) in Table 6.1 with theoretical ones. We can see that the measured time and speedup factor are close to the theoretical values. Since more nodes are involved in performing the 3D reconstruction job, larger data access bandwidth and more computing resources contribute to the nearly linear increase of reconstruction speed. For each node in the cluster, data are accessed via 1 GE and processed by the dual quad-core processors. As more nodes are involved, the data access bandwidth and computing power proportionally increase, leading to the great decrease in time and increase in speedup factor.

In the second stage of measurement, multiple parallelized mapping tasks are handled by every node to better utilize its computing power. Table 6.2 shows the the recorded time and speedup factors in the stage of multiple mapping task per node, where the number of mapping tasks is equal to the number of mapping tasks per node times 37 (the number of nodes). As the number of mapping tasks per node increases from one to six, the time required by the 3D reconstruction job decreases from 1107s to 330s, while the speedup factor goes up from 35.85 to 120.0. Their comparisons with the theoretical values of
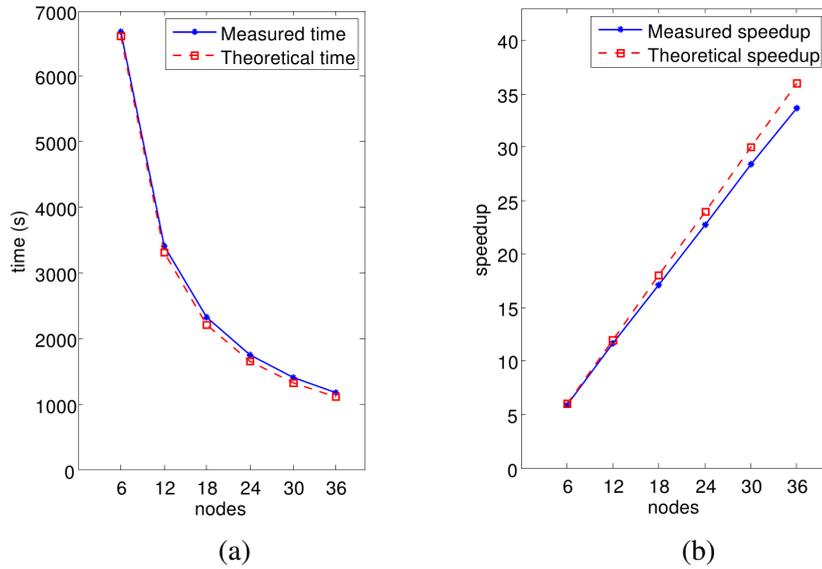
Fig. 6.3. Parallel computing performance of 3D reconstruction in the stage of one mapping task per node: (a) time in seconds; (b) speedup factor.

reconstruction time and speedup factor are plotted in Fig. 6.4. We can see that the distance between the measured lines and the theoretical counterparts becomes larger with the increase of mapping tasks per node.

Compared with the performance in the first measurement stage shown in Fig. 6.3, larger distance between the measured and theoretical curves are displayed in Fig. 6.4. This is mainly due to the increasingly limited computing resources in each node, such as memory and processing cores, in executing more mapping tasks simultaneously. Since each mapping task has been multithreaded in MATLAB R2010b by default, multithreads can run on different cores and affect other tasks. In the case that the computing resources are not sufficient, some part of these tasks have to be suspended until the computing resources are released, which is determined by the operating system. In this way the parallelized mapping tasks share the same computing resources in each node and compete with each other. As a result, slower growth of computing performance is displayed compared with the first measurement stage. However, a speedup factor 120.0 is achieved if the maximum number of mapping tasks per node (6) are included in each node. A time (330s, less than 6 minutes) are actually required for the whole 3D reconstruction job, suggesting that the presented parallel reconstruction framework with LSDF is able to handle the 3D tomographic reconstruction in near real time.

Table 6.2. Computing performance in the stage of multiple mapping tasks per node

| mapping tasks per node	mapping tasks	time (s)	speedup factor
1	37	1107	35.85
2	74	650	61.08
3	111	488	81.39
4	148	417	95.20
5	185	360	110.2
6	**222**	**330**	**120.0**
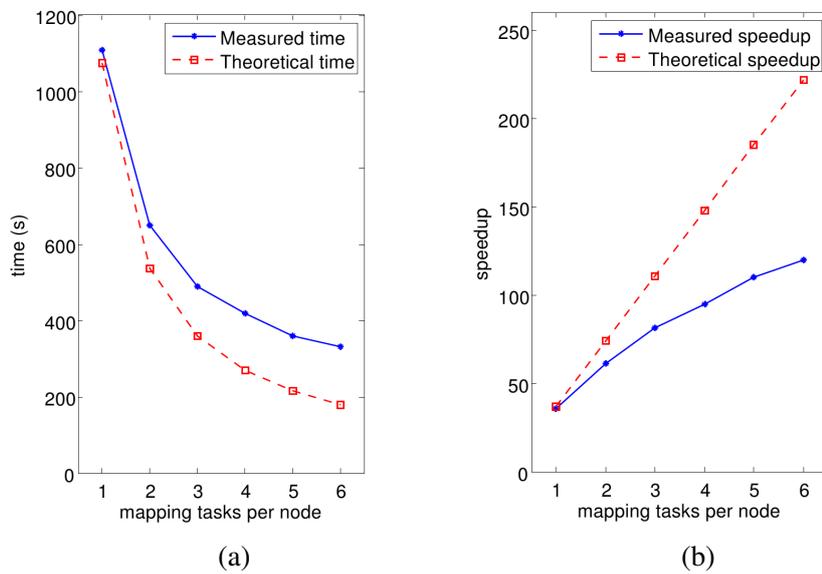
Fig. 6.4. Parallel computing performance of 3D reconstruction in the stage of multiple mapping tasks per node: (a) time in seconds; (b) speedup factor.

6.5. Summary

In this chapter, the intensive computing of iterative reconstruction with tomographic data from ANKA is studied by employing the compute cluster at the large scale data facility (LSDF). The LSDF at KIT is designed to support multiple disciplines and institutes such as the tomography beamline of ANKA. It offers not only nearly unlimited storage, but also additional services of data processing, analysis and access for the users. An automated workflow is built up to connect the tomography beamline of ANKA with LSDF, enhancing the data storage and analysis efficiency.

The application of iterative reconstruction is motivated by the requirements of fast tomographic imaging and low X-ray radiation dose. It is able to produce high quality reconstruction image even a limited number of projections exists in the sense of FBP reconstruction. However its intensive computing property makes the 3D reconstruction job very time-consuming. The parallel computing framework using the computer cluster at LSDF in this chapter enables its practical application. Data parallel computing of 3D CT reconstruction is performed by distributing the reconstruction job into a number of parallel tasks. Compared with the sequential reconstruction taking more than ten hours, the parallel computing using 37 nodes at the LSDF only needs less than 6 minutes with the speedup factor reaching up to 120. This near real time performance satisfies the requirements of offline analysis for large tomographic data. Moreover, a great advantage of this parallel reconstruction framework is its good scalability in processing large datasets. Namely, if more cluster nodes are used, better time performance can be achieved.

7. Discussion

In this thesis an improved method for precise and automated tomographic reconstruction with a limited number of projections is presented. The application of this method is motivated by the requirements of reducing the X-ray radiation dose and adapting to the fast tomography setup in the 4D *in vivo* tomographic imaging in order to investigate tissue development and organ movement of a biological object. The characteristics of the presented reconstruction method are threefold. Firstly, it is able to produce high quality image even using insufficient projections. Secondly, the implementation of the reconstruction algorithm can be reliably carried out in an automated way without significant manual intervention. Thirdly, the reconstruction algorithm can be implemented in a short time (in the order of 10 min) acceptable for offline data analysis. The presented reconstruction method completes the workflow of data management and processing designed for tomographic experiments. The application of the Large Scale Data Facility (LSDF) makes the automated implementation feasible and applicable.

Since the FBP reconstruction method is sensitive to the decreasing number of projections and leads to serious streak-line artifacts, this thesis resorts to the iterative techniques which are effective on truncation artifacts. However, the iterative reconstruction method is challenging due to the parameter determination as well as its intensive computing. In order to simplify the parameter problem, this thesis keeps the minimization problem for reconstruction in a simple form which includes only one key parameter λ controlling the trade-off between the data fidelity and the regularization terms. The value of λ is chosen by the scientist based on experience commonly. However, in practice the optimal value of the parameter is changeable for different datasets because of various noise level, sparsity, and gray scale. The parameter needs to be adapted to dataset-specific conditions. For example, a dataset with lower level of noise and sparser representation tends to have an increased value for the optimal parameter λ. Therefore, the manual determination of the parameter's value is an annoying operation and prevents the realization of an automated reconstruction process.

In this thesis, a simple implementation of the conjugate gradient method adapted to the TV regularization is applied to directly solve the minimization problem with only one yet undetermined parameter λ. The determination of a proper value of the parameter is automatically performed by the L-curve method, which guarantees an approximately optimal parameter value with respect to dataset-specific conditions, and thus obtains a reasonable

and robust reconstruction result. Consequently, a parameter-optimized, automated iterative reconstruction method for computed tomography, named as the optimized CGTV method, is developed. The expensive computing required by the iterative reconstruction is greatly alleviated by parallel slice reconstruction through a computer cluster which is directly connected to the LSDF. The parallel reconstruction framework accelerates a 3D reconstruction of volumetric data by over hundred times compared to the sequential implementation.

The proposed method has been evaluated in detail by investigating the reconstruction image quality for several datasets, including a simple computer-simulated phantom, a generic image containing more complicated structures, and real experimental datasets acquired by the propagation-based X-ray *in vivo* imaging of biological objects. From the evaluation we can conclude that the proposed method represents a promising venue for reconstruction with limited projections. The reconstruction result maintains resolution in an acceptable level, reproduces edges bounding sufficiently large structures, quickly saturates the image quality as the number of projections increases, and is void of streak-line artifacts as generated by the filtered back-projection (FBP) method. This is important for automated image analysis such as structure segmentation using a priori set of gray-value thresholds. With the parallel reconstruction framework in the computer cluster, the actual time demanded by the 3D volume reconstruction is in the order of 10 minutes.

In the following sections the proposed reconstruction method as well as the obtained results will be discussed.

7.1. Choice of the iterative reconstruction method

The computed tomography reconstruction methods are generally divided into the analytical reconstruction group and the iterative reconstruction group. The FBP is taken as the representative method for analytical reconstruction. Overall speaking, the main advantages of the FBP include the reconstruction speed, high-quality and robust reconstruction in the case of sufficient projection data. FBP-based algorithms are still the most widespread implementations of the CT reconstruction technique, especially in the area of commercial CT manufacture [130]. Its main disadvantage is the unacceptable reconstruction image quality due to the noise and artifacts if limited projection data are available. More and more explorations of the use of IR are made in scientific research, such as the biological and medical study. There new practical requirements like reducing the X-ray radiation dose come up. In the early age of computed tomography, the application of IR was limited by its high demands on computing power. Also, most algorithms exist in the literature are theoretically working with small artificial data. Real world data are always challenging. With the rapid development of computer technology as well as the reconstruction technique, the usage of IR is becoming a more realistic, popular option in

consideration of its greatly improved image quality over the analytical FBP reconstruction in the context of reducing the number of projections. In order to choose an appropriate reconstruction method for a practical application, one must take into account all these factors including the reconstruction image quality, the scale of the problem, the computational requirements, robustness and stability.

As the problem this thesis focuses on is the tomographic reconstruction from a limited number of projections in order to reduce the radiation dose or adapt to the imaging setup. Namely, the projections are far from sufficient to reconstruct a high quality object image in the sense of FBP reconstruction. In this context the iterative reconstruction methods are preferred in order to produce high quality reconstruction image with the help of prior information and regularization method. They tend to eliminate streak artifacts and provide a more accurate body contour. However, one has to notice that the improvement with the iterative reconstruction is obtained at the expense of increased calculation time. With the present computer technique it is practical to perform the reconstruction of a general two-dimensional image. Thus for the parallel imaging geometry a three-dimensional reconstruction image are often obtained by stacking two-dimensional reconstruction results. A real three-dimensional reconstruction is still challenging due to the big system matrix and a large amount of matrix multiplications. Furthermore, the iterative reconstruction is more robust to the variances in the projection data such as the level of noise and the detector defects. To sum up, the choice of the iterative reconstruction for our purpose is a good balance between the reconstruction accuracy and computational efficiency.

7.2. Precise and automated iterative reconstruction

Key parameter

In iterative reconstruction methods, parameters are often involved in solving the numerical minimization problem. They are either given with an fixed value or adapted within iterations according to the newly updated result. The importance of the parameter is defined by checking if the parameter plays an important role in the iterative process and has a big influence on the reconstruction result. The iterative process as well as the reconstruction result could be robust to the less important parameters, for which little effort can be made to determine their values and some roughly selected values are sufficient. More important parameters need to be determined by giving more precise estimation to their values or adapting them according to the intermediate result in current iteration.

In the regularized minimization problem for reconstruction, the regularization parameter λ is the most important parameter, which controls the trade-off between the data fidelity

term and the regularization term. The minimization of the data fidelity term ensures the reconstructed result is well fit to the measured projections, and the minimization of the regularization term produces a smoothness required by the sparsity assumption in the total variation. In general, an image is comprised of multiple features at different scales. Roughly speaking, the image with large features tends to be much sparser, for which a large λ leads to smoothing so that noise is removed considerably. On the other hand, the image with small features tends to be less sparser, for which a small λ leads to little smoothing and usually good detail preservation. The well-determined value for the regularization parameter λ, fitting to the features of interest in the object image, is supposed to improve the reconstruction image quality.

Nevertheless, there are some advanced solvers to solve the minimization problem in the literature in order to achieve fast convergence property, where the parameter λ disappears but one or more alternative parameters are actually used in the reconstruction process. This increases the complexity of parameter determination. Even though these parameters still play the similar roles, it is not clear to see the interactions and relationships between the parameter values and the reconstruction result. Especially high-dimensional parameter determination is a very challenging task. In order to keep the parameter determination as simple as possible, the solvers which solve the minimization problem directly are preferred, this thesis focused on the determination of the regularization parameter in the minimization problem.

Parameter determination

To determine the regularization parameter λ in the minimization problem, the method based on L-curve was used. The curve is constructed by giving different values to the parameter λ and then implementing the reconstruction algorithm with a fixed parameter. By reviewing the L-curves obtained for different datasets, we can clearly see that the curves constructed are shaped as the letter 'L' which includes a vertical curve and a horizontal curve. The vertical curve is related to small λs, where the reconstruction result normally obtains a good data fidelity (small value for the term $\|\mathbf{Ax} - \mathbf{p}\|_2^2$). It tends to preserve small features in the image, but meanwhile is sensitive to the noise in the projection data and produces artifacts. On the other hand, the horizontal curve represents the reconstructions with large λs, which highlight the regularization term so that the reconstruction result obtains a smoothing effect. It is robust to the noise in the projection data and reduces artifacts, but tend to remove small features in the image. A good determination for the parameter λ is the 'corner' of the whole curve, where the reconstruction result is considered to reach a good trade-off between data fidelity and noise and artifacts reduction.

Regarding the approach to locate the L-curve corner, the common idea is to search the point at which the curve has the maximal curvature. Practically, it is not possible to

construct the L-curve for all λ values. Thus only a discrete set of values were used to determine the location of the real corner. The discrete set of λs produce a set of points in the curve. Namely, only a small set of points in the curve are known, which brings difficulties in the determination of the corner. The real corner of the curve could locate at one of these points or between two points. The criterion of maximal curvature requires complex curve fitting and interpolation problem. This makes it not necessary to exactly locate the position of the real corner. The approach in this thesis intended to determine an approximate position of the corner, thus the criterion of minimal distance to the origin was applied and one of the set of points was selected as the corner.

The L-curve method is a popular practical method for the regularization parameter determination. Its main limitation is that its computation is potentially an expensive task considering that the solving process of the regularization problem should be repeated for many values of the regularization parameter. The simple way of determining the corner of the L-curve in this thesis reduced the requirements of computation. For further applications which demand more accurate value for the regularization parameter, a method which can iteratively track the curve property in a small range nearby the real corner and determine an 'optimal' point should be used. It starts from an initial set of discrete points in the curve and will require more points nearby the optimal selection. This method increase the accuracy of the selected value for λ, but generally at the expense of higher computation demand.

Automated reconstruction strategy

Once a good value for the parameter λ was fixed, the regularized minimization problem was solved by the nonlinear conjugate gradient method, which was referred to as the optimized CGTV method. As the 3D reconstruction of the projection data acquired in a parallel beam is comprised of a series of 2D slice reconstructions, strictly speaking, it is necessary to separately determine the best value for each slice. Even though this slice-specific parameter determination and reconstruction can be performed in a parallel way, it still lacks of computing efficiency. In this thesis a better strategy is proposed, which determines the value of λ for a typical slice and subsequently uses it for the reconstruction of all other slices. It ignores the possible difference in sparsity of different slices and its affects on the optimal value for λ. Nevertheless, this treatment still makes a lot of sense as it is able to preserve the features in the same scale in the 3D reconstruction result. Thus the parameter determination with the L-curve method enabled an automated reconstruction process and guaranteed a relatively reliable reconstruction result.

7.3. Reconstruction in data management workflow

The optimized reconstruction method was embedded in a data management workflow based on the Large Scale Data Facility (LSDF), which aimed to connect the tomographic experiment station to the LSDF in order to enhance the data storage, processing, and analysis for large volume of tomographic data acquired from the tomographic experiment. The workflow illustrates the data flow starting from the data acquisition to the finial results required by the user. Its realization requires contributions from multiple scientific research communities. The reconstruction method in this thesis comprises one component of this workflow, which is a key step for analyzing and understanding the actual information included in the original datasets.

The reconstruction process should be invoked and executed automatically in the workflow, which was guaranteed from two points. On the one hand, from the point of view of reconstruction algorithm, the parameter determination with the L-curve method enabled an automated reconstruction without manual interference and produced a promising reconstruction result. On the other hand, the workflow provided an automated strategy for reconstructing the tomographic data from the point of view of software system and hardware architectures. It applies the capabilities of LSDF for large storage and intensive computing. The computing power is offered by the connected computer cluster where the reconstruction job is implemented in parallel to accelerate the reconstruction process. The parallel reconstruction framework was based on the geometric property of the parallel X-ray beam, which divided the whole reconstruction job into a series of independent slice reconstruction tasks distributed in different computer nodes. The assumption of the independent slices has limitation to the reconstruction result, that is the asymmetry of the reconstructed 3D data in different dimensions, the vertical direction versus the slice plane. The parallel computing framework was successfully tested independent of the reconstruction method, which obtained an accelerated reconstruction process by a factor of 120. It can be regarded as a real time reconstruction in the case of offline analysis.

The reconstruction method was implemented in MATLAB. It is a common programming method in the filed of digital image processing due to the convenient vector/matrix operations and developed toolboxes providing a large amount of functions for signal/image processing. It also contributes the aim of producing an extendable software. The scripting language based programming enables the integration of new functions and also the extension to the existing functions. Thus MATLAB allows a programming environment for rapid algorithm development, data analysis, and visualization. Besides, scripts written in MATLAB can be executed by using the command lime, allowing their remote execution easily. Therefore, MATLAB was used for implementing the algorithms in this thesis. The usage of MATLAB may increase calculation time, which is expected to be relived by the parallel computing framework in the workflow. Further improvement in the computing

performance can obtained by code optimization and the use of GPU acceleration specially for the forward imaging operation which is computationally expensive and carried out numerous times. This will require more work of programming in other programming languages and on other architecture.

7.4. Reconstruction results

The optimized reconstruction method proposed in this thesis was evaluated with several datasets, including computer simulated images and real experimental *in vivo* data acquired by propagation based X-ray imaging. The computer simulated images are the sparse Shepp-logan phantom and a relatively more general image which contains a lot of structures and details. The experimental data are tomographic imaging results of a living weevil and a developing frog embryo. Even though the optimized reconstruction method is also tested with other more datasets, the results are not shown in this thesis since they illustrated similar conclusion. Different datasets shown in this thesis are regarded as representative ones. The raw data acquired from practical experiments generally contain several hundred to over a thousand projections for achieving promising reconstruction result from FBP reconstruction method. In order to evaluate the optimized reconstruction method, a small number (~ 100) of projections were extracted evenly.

Regarding the reconstruction image quality assessment, several influencing factors have to be considered. The first problem is the absence of ground truth for real word datasets. Artificial datasets do not have this problem as the true original images exist. For the experimental datasets, it is common to approximately take the FBP reconstruction image obtained from all available projections in the raw data as the reference image. Another problem we need to consider is the quantitative measures which can objectively reflect the properties of the reconstruction image. The evaluation by visually observation by human eyes gives straightforward concept of the image quality, but is limited. Objective evaluation with quantitative measures can verify the subjective judgment. As the structures and details are the key features comprising the main framework and contents in the image, the image quality assessment should take into account the structural information besides the gray level based evaluation method. Thus in this thesis, two measures were used: the mean square error (MSE) and the structural similarity (SSIM). The former one compares the gray value of corresponding pixels in the reconstruction and reference image with assumption that the image pixels are statistically independent. In contrast, the SSIM integrates the structure comparison and thus the dependency of image pixels is also taken into account. The simultaneous usage of them enables a comprehensive assessment.

High quality reconstruction results are obtained by using the optimized reconstruction method. By the visual evaluation, the reconstruction result by optimized CGTV is void of

streakline artifacts as generated by FBP reconstruction due to limited number of projections. On the other hand, the L-curve method determined a proper value for the trade-off parameter λ, which produces such a promising reconstruction as other advanced implementations with total variation regularization (such as TVAL3 and TwIST). It maintains resolution at an acceptable level and realistically produce edges which bound sufficiently large structures such as cells and tissue confines. The promising image quality of the optimized CGTV reconstruction was also demonstrated by the quantitative measures (MSE and SSIM). All the reconstructed images were compared to the corresponding reference image. The quantitative evaluation showed a great improvement in the reconstruction by the optimized CGTV compared to the FBP reconstruction in the case of limited number of projections. And the reconstruction results by the optimized CGTV were not worse than other TV-based implementations (TVAL3 and TwIST). The high quality reconstruction of the optimized reconstruction is attributed to on the one hand the usage of the regularization of total variation for sparsity-based reconstruction and on the other hand the optimized parameter determination for the balance of artifact reduction and important detail preservation.

The high reconstruction quality is important, e.g., for automated image analysis such as structure segmentation, according to a prior set grey-value thresholds. That is, moderately sacrificing the spatial resolution is not a strong limitation if the subject of biological research is to understand coarse phenotypic dynamics such as tissue separation, cell division and migration, and cavity formation. All these only require a clear segmentation of the associated boundaries. Because of this,the optimised CGTV should be interesting for 4D livecell imaging, aspiring to the acquisition of long time-lapse series [14, 15]. The *in vivo* data, as reconstructed in the present work, is not contaminated by a high level of statistical photon noise. Exemplarily, we show for the weevil data that with an artificial increase of Poisson noise optimal CGTV performs well in annihilating noise-induced small-scale fluctuations at the same time well preserving objects features seen in CGTV reconstruction based on the low-noise data. A more systematic investigation of optimal CGTV reconstruction under various sources of noise is certainly in order. We plan to carry out such an analysis in future work based on *in vivo* tomographic data.

The TV regularization based tomographic reconstruction highly relies on the sparsity assumption of the object to restore. The sparser is the object, the better is the reconstruction image for a fixed number of projections. In other words, the number of projections required for a valuable reconstruction is related to the sparsity of the object. This can be seen from the reconstruction of the Shepp-Logan phantom versus the Barbara image and the weevil dataset versus the frog embryo dataset. The fine details violate the sparsity of the image. The phantom is very sparse as it contains only several relatively large structures. In contrast, the Barbara image contains more fine details. Thus 120 projections are actually used for the Barbara image instead of 60 projections for the Shepp-Logan phantom. Similarly, the sparsity assumption of the gradient image for TV regularization is held well for the weevil dataset. In contrast, the frog embryo contains a large number

of cells which lead to some level of variation within regions. Thus 167 projections were used for the frog embryo instead of 80 projections for the weevil dataset. However, all of them are still in the range of limited projections where the FBP reconstruction is not able to generate a satisfied result. The optimized reconstruction quickly saturates the image quality as the number of projections increases. The FBP reach a stable reconstruction only when the number of projections goes to a large value.

One may note that a phase retrieval process might be involved before the reconstruction of experimental datasets in this thesis. Strictly speaking, the gray value reconstructed from the projections does not mean exactly the attenuation coefficient since the phase contrast was included in the projection images in order to increase the contrast of different biological tissues which are generally difficult to be distinguished by only attenuation effect. For the weevil dataset the attenuation effect mainly contributes to the final reconstruction contrast which, however, is somehow enhanced by the phase contrast without phase retrieval. In contrast, the frog embryo contains only soft tissues which can not be distinguished by attenuation effect. The phase retrieval of the projections based on phase-contrast imaging technique can provide differences and contrast of soft tissues. Due to the existing of phase contrast, the reconstructed gray value might be negative, which does not conform to its physical meaning any more. However, the enhanced contrast is indeed helpful for the object understanding and analysis.

The reconstruction is a significant step in the whole scientific workflow. Apart from the reconstruction image quality, the automated property and fast reconstruction speed are also in the concern for practical purpose.

The automated property of the optimized reconstruction in this thesis was mainly attributed to the usage of an automated parameter determination method based on L-curve. It provides not only a proper value for the regularization parameter to guarantee a valuable reconstruction but more importantly an automated venue which is void of trivial manual operation based on professional experience. Note that the regularization parameter λ determined by the shortest distance to the origin in the discrete L-curve may not be the optimal one. More accurate determination can be pursued by using more complex criterion such as the maximum curvature of a fitted curve. The L-curve can not be guaranteed to be always perfectly 'L'-shaped. It can be influenced by the instability of the iterative reconstruction related to the selected reconstruction algorithm, the initial guess, the number of iterations, and the noise level of the raw data.

Regarding the reconstruction speed, the parallel reconstruction framework improved the computing performance of a 3D reconstruction for a volume data by hundredfold, which can satisfy the requirements for the offline analysis in the data management workflow designed in this thesis. However, it is still not comparable to the real-time FBP reconstruction normally implemented by using GPU infrastructures, which is often applied by online analysis system. The near real-time property of the optimized reconstruction for a volume data was mainly attributed to the parallel reconstruction in the data level. In the

future work, the computing performance can be further improved by integrating the reconstruction method with faster convergence rate in the algorithm level. Namely, instead of the conjugate gradient method, certain accelerated strategy in the reconstruction algorithm can be applied, such as the multistep version of an gradient-like method [131].

7.5. Recommendations for the future

A further improvement of the reconstruction method would be the application of a more accurate forward model, which plays an important role in building up a more consistent equation system for the algebraic reconstruction technique. One advantage of iterative reconstruction method is its convenience to incorporate information in the forward model, such as the geometric configuration of the sensor and the defect sensor elements. The more precise is the model, the less artifacts we expect in the reconstruction. The challenge is the complexity of calculating the system matrix. In the practical implementation of the reconstruction in this thesis, the X-ray beam is modeled as a 'thin' straight line going through the imaged object, which is a compromise between the demanding computing time and reconstruction preciseness. Considering that the creation of the system matrix can be decomposed to independent calculations of the contribution of all pixels to each projection, the forward process can be easily implemented in a parallel way. As the forward process is the most often used operation in the whole reconstruction procedure, we save the system matrix in the storage so that it can be calculated once for accessing many times. However, accessing to the large matrix in a particular storage might take longer time than its calculation when applying parallel implementation. Hence, the parallel implementation of the forward process will further accelerate the whole reconstruction procedure without saving the large system matrix in the storage, consequently enabling the application of accurate forward models.

As the problem of reconstructing the object image from limited number of tomographic projections is typically ill-posed, stable reconstruction process highly relies on regularization techniques which are based on prior information of the unknown image. The total variation applied in this thesis is a typical regularization in this direction, which extracts the sparsity of the image in the gradient domain aiming to preserve significant edges in images. In this thesis, only a general expression of the total variation was used. But if we carefully observe a homogeneous region, there are often small pieces with slightly varying gray values forming staircase artifacts. This piece-wise consistence results in unnatural impression in understanding structures in the reconstruction image. It can be improved by applying the extended versions of total variation in the literature [44, 108, 132, 133], all of which aim to preserve even finer edges while smoothing artifacts and noises by considering the multi-scale property of the image features, the mixed norm of directional image derivatives, or the penalty weights with respect to the anisotropic edge property.

Apart from total variation, more sparse representations of the image can be also taken into account with expectation of improving reconstruction result, such as the wavelet and dictionary learning based regularization [92, 90].

Due to the parallel-beam illumination, this thesis only focused on the reconstruction of one 2D slice. The reconstruction of 3D volume was accomplished by vertically stacking such 2D slices, which contributes to the parallel computing structure for fast reconstruction. However, this produces an asymmetric treatment of sparsity in x, y (slice plane) versus z directions. Namely, the sparsity of the gradients in only x and y directions are integrated into the 2D total variation without taking into account the variance in z direction. No matter the regularization parameter λ takes either slice-specific value or a common value derived from a typical slice, the vertical stacking of 2D slices can create discontinuity of the value in z direction. This can be seen from the slight jumping of the gray value in x-z and y-z plane even inside the same structure. Thus in z direction the features in the object are not well reconstructed.

This can be improved by considering 'real' 3D reconstruction. By 'real' 3D reconstruction we mean that the 3D volume is reconstructed directly from the reconstruction algorithm using all projection data in z direction instead of vertical stacking of 2D slices. In the 'real' 3D reconstruction, the pixel will be changed to voxel, which requires new modeling of the X-ray beam, forward imaging process, and the total variation. In 3D case, the amount of voxels is much greater than the pixels in 2D reconstruction. Thus the tomographic reconstruction becomes the inverse problem of an even bigger equation system as well as the resulting optimization problem. The challenges in solving this big problem include the calculation of the large system matrix and expensive computing of the reconstruction algorithm. For a typical reconstruction of a 2D image 1024×1024, the system matrix takes up ~ 3GB storage due to its sparse property. In contrast, for the reconstruction of a volumetric image $1024 \times 1024 \times 1024$, the size of the system matrix goes up to ~ 3PB, which is challenging in storage and calculation. Due to lack of data parallel reconstruction structure, the 3D reconstruction is almost an impossible task using present computing infrastructure. But it is still possible to search solution by parallel computing at the algorithm level combined with the high performance computing infrastructures. In the near future, the difficulties will be definitely overcome by the fast developing of computer techniques. The 3D even as well as 4D reconstruction can be practical in general computing architecture. The segmentation and reconstruction can be completed in one regularization step.

7. Discussion

8. Conclusion

This thesis presents a robust and automated tomographic reconstruction for the LSDF-based data management workflow by proposing a parameter-optimized algebraic reconstruction method for computed tomography with a limited number of projections. It is subject to the minimization of total variation in the sense of compressive sampling theory to pursue a precise and automated 3D reconstruction. The optimized CGTV method, on the one hand, incorporates the regularization of total variation to accommodate such prior knowledge as the existence of sparse representation of original object. On the other hand, it uses a parameter-optimized method of L-curve in fixing the value of important trade-off parameter in an automated way. Thus it provides high quality reconstruction results and an automated implementation of 3D reconstruction without manual adjusting operations. A simple implementation of the conjugate gradient method adapted to the TV regularization is used to solve the unconstrained reconstruction problem.

The evaluation of the the optimized CGTV with the simulated images and *in vivo* data acquired by propagation based X-ray imaging demonstrates that the method represents a promising venue for reconstruction with limited projections. It maintains resolution in an acceptable level, reproduces edges bounding sufficiently large structures, quickly saturates the image quality as the number of projections increases, and is void of streakline artifacts as generated by the filtered back-projection (FBP) method. This is important for automated image analysis such as structure segmentation appealing to a priori set of gray-value thresholds. The expensive computing required by the algebraic reconstruction of a full 3D volume is greatly alleviated by parallel volume reconstruction through a computer cluster connected to the Large Scale Data Facility (LSDF).

The application of the optimized tomographic reconstruction enables the construction of a fast tomography beamline at ANKA. Applying the iterative reconstruction method reduces the number of projections from $\sim$1000 to $\sim$100 without significant loss in the reconstruction image quality, thus also reduces the data acquisition time by over tenfold compared to the conventional method. By the fast tomography beamline the 3D imaging procedure takes such a short snap that the object movements can approximately be omitted with little influence on the reconstruction. This enables biological scientists to precisely investigate the movements, organ structures, and dynamic characteristics by 4D *in vivo* imaging of a biological object, like the weevil used in this thesis.

Also the application of the optimized tomographic reconstruction reduces the X-ray radiation dose for biological object. Thus it increases the duration of *in vivo* time-lapse series for the developmental frog embryo, allowing biologists to observe the embryo's developing procedure for longer time and track the dynamic characteristics of cells or issues.

Apart from the biological applications, the presented reconstruction method apparently can be also used in other fields, such as nanostructured material analysis by electron tomography where also a limited number of projections is available shown in Chapter 4.

The proposed reconstruction method provides a promising venue for fast tomography and low-dose tomographic imaging. It therefore is likely to direct biologists' plan in experimental design, data acquisition and analysis. It may also contribute to the construction of a fast tomography beamline.

Appendix

A. Imaging information for experimental datasets

Weevil dataset

For imaging the weevil, white-beam illumination with a critical energy of $E_c \sim 15\,\text{keV}$, a propagation distance of $z = 50\,\text{cm}$, a photon flux density of $\sim 10^{13}$ photons/mm^2/s, an effective pixel size of $\Delta x = 3.7$ µm, and a field of view of $(7.4\,\text{mm})^2$, corresponding to 2000×2000 pixels, was employed at ANKA's TOPO-TOMO bending-magnet imaging beamline. The detection system uses a freestanding 50 µm-thick Cerium-doped lutetium aluminum garnet (LuAG:Ce) scintillator orthosilicate scintillator, generating visible light by the absorption of X-rays. This latent visible-light image is relayed by an optical system with a three-fold magnification onto the pco.dimax camera that performs the actual signal detection. An acquisition of 400 intensity projections was performed per tomogram with an exposure time of 0.3 ms per projection.

Frog embryo dataset

The frog embryo was imaged at the undulator imaging beamline 32-ID of APS subject to monochromatic ($\Delta E/E \sim 10^{-4}$), highly coherent X-ray illumination of energy $E = 30\,\text{keV}$, a photon flux density of $\sim 10^{12}$ photons/mm^2/s, a propagation distance of $z = 70\,\text{cm}$, an effective pixel size of $\Delta x = 1.3$ µm, and a field of view of $3.328 \times 2.808\,\text{mm}^2$, corresponding to 2560×2160 pixels. The detection system used in this experiment was a 100 µm LuAG:Ce scintillator supplemented by a five-fold magnifying optics and a pco.edge camera. The number of projections acquired per tomogram was 499 with 60 ms of exposure time per projection. A rough estimate yields that the dose per projection, deposited into the object, is about 20 times smaller for weevil compared to frog-embryo imaging.

A. Imaging information for experimental datasets

B. Hardware and software setup for the LSDF

Hardware

The LSDF hardware infrastructure provides a computer cluster of 58 Hp ProLiant DL1000 nodes with dual quad-core Intel processors and 36 GB memory to allow data intensive computing close to the data. For external data access, each institute using the LSDF is connected via 10 Giga Ethernet (GE), for internal access each cluster node can access data via 1 GE. To offer flexibility to a large extent, these resources can be operated either as a Cloud infrastructure or as a Hadoop [134] computer cluster. Among these 58 nodes 21 are set to provide services for Cloud computing, and other 37 nodes are utilized for the MapReduce-enabled analysis [135] provided by the Hadoop framework. The LSDF Hadoop installation offers the MapReduce job submission component and a HDFS filesystem comprising 110 TB disk capacity. For direct user access the Hadoop filesystem is made available via a FUSE [136] mount. For access to large scale data each cluster nodes has access to a shared filesystem where the LSDF scientific data is stored. This is a GPFS [137] filesystem running on high-end storage hardware (DDN SFA10000), and mounted from all the processing nodes via the NFS protocol.

Software

All compute cluster nodes are running the Scientific Linux 5 [138] operating system. In addition to the base system, almost arbitrary software components can dynamically be configured from the runtime environment, using the Environment Modules [139] tool. This tool is used to define customized environment variables and scripts for each software component deployed in the cluster to enable/disable it dynamically. The user, or jobs by the user, only has to load the required modules by executing simple commands. Thereby, each user and job gets a clean runtime environment and results are reproducible at all time. Other advantages are that different versions of the same software package can be supported at the same time and that all supported software packages are centrally managed as part of the cluster administration and not by the users themselves. One of these

software packages relevant for this thesis is MATLAB. MATLAB offers a programming environment for rapid algorithm development, data analysis and visualization. Apart from this, scripts written in MATLAB can be executed on the command line, which allows their remote execution easily. Due to many contained additional packages and the campus license available at KIT MATLAB is a perfect candidate for implementing algorithms needed for this work.

Bibliography

[1] A. G. Filler. The history, development and impact of computed imaging in neurological diagnosis and neurosurgery: CT, MRI, and DTI. Nature Precedings, 7(1): 1–69, (2009).

[2] J. Radon. 1.1 Über die Bestimmung von Funktionen durch ihre Integralwerte längs gewisser Mannigfaltigkeiten. Classic papers in modern diagnostic radiology, 5, (2005).

[3] F. Peeters, Verbeeten J. B., and H. W. Venema. [Nobel Prize for medicine and physiology 1979 for AM Cormack and GN Hounsfield]. Nederlands tijdschrift voor geneeskunde, 123(51):2192–2193, (1979).

[4] G. T. Herman. Fundamentals of computerized tomography: image reconstruction from projections. Springer Science & Business Media, (2009).

[5] B. Unal, A. Aktas, G. Kemal, Y. Bilgili, S. Guliter, C. Daphan, and K. Aydinuraz. Superior mesenteric artery syndrome: CT and ultrasonography findings. Diagnostic and Interventional Radiology, 11(2):90–5, (2005).

[6] R. C. Krempien, S. Daeuber, F. W. Hensley, M. Wannenmacher, and W. Harms. Image fusion of CT and MRI data enables improved target volume definition in 3D-brachytherapy treatment planning. Brachytherapy, 2(3):164–171, (2003).

[7] J. Cai, J. C. H. Chu, D. Recine, M. Sharma, C. Nguyen, R. Rodebaugh, V. A. Saxena, and A. Ali. CT and PET lung image registration and fusion in radiotherapy treatment planning using the chamfer-matching method. International Journal of Radiation Oncology Biology Physics, 43(4):883–891, (1999).

[8] O. Gaemperli, T. Schepis, I. Valenta, L. Husmann, H. Scheffel, V. Duerst, F. R. Eberli, T. F. Luscher, H. Alkadhi, and P. A. Kaufmann. Cardiac image fusion from stand-alone SPECT and CT: clinical experience. Journal of Nuclear Medicine, 48 (5):696–703, (2007).

[9] E. Nkenke, S. Zachow, M. Benz, T. Maier, K. Veit, M. Kramer, S. Benz, G. Häusler, F. W. Neukam, and M. Lell. Fusion of computed tomography data and optical 3D images of the dentition for streak artefact correction in the simulation of orthognathic surgery. Dentomaxillofacial Radiology, 33(4):226–232, (2004).

[10] G. Li, D. Citrin, K. Camphausen, B. Mueller, C. Burman, B. Mychalczak, R. W. Miller, and Y. Song. Advances in 4D medical imaging and 4D radiation therapy. Technology in cancer research & treatment, 7(1):67–81, (2008).

[11] T. van de Kamp, P. Vagovič, T. Baumbach, and A. Riedel. A biological screw in a beetle's leg. Science, 333(6038):52–52, (2011).

[12] T. dos Santos Rolo, A. Ershov, T. van de Kamp, and T. Baumbach. In vivo X-ray cine-tomography for tracking morphological dynamics. Proceedings of the National Academy of Sciences, 111(11):3921–3926, (2014).

[13] A. Rack, T. Weitkamp, S. Bauer Trabelsi, P. Modregger, A. Cecilia, T. dos Santos Rolo, T. Rack, D. Haas, R. Simon, R. Heldele, et al. The micro-imaging station of the TopoTomo beamline at the ANKA synchrotron light source. Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms, 267(11):1978–1988, (2009).

[14] J. Moosmann, A. Ershov, V. Altapova, T. Baumbach, M. S. Prasad, C. LaBonne, X. Xiao, J. Kashef, and R. Hofmann. X-ray phase-contrast in vivo microtomography probes new aspects of Xenopus gastrulation. Nature, 497(7449):374–377, (2013).

[15] J. Moosmann, A. Ershov, V. Weinhardt, T. Baumbach, M. S. Prasad, C. LaBonne, X. Xiao, J. Kashef, and R. Hofmann. Time-lapse X-ray phase-contrast microtomography for in vivo imaging and analysis of morphogenesis. Nature protocols, 9 (2):294–304, (2014).

[16] J. Moosmann, R. Hofmann, and T. Baumbach. Single-distance phase retrieval at large phase shifts. Optics Express, 19(13):12066–12073, (2011).

[17] R. Hofmann, J. Moosmann, and T. Baumbach. Criticality in single-distance phase retrieval. Optics Express, 19(27):25881–25890, (2011).

[18] S. Do, W. C. Karl, S. Singh, M. Kalra, T. Brady, E. Shin, and H. Pien. High Fidelity System Modeling for High Quality Image Reconstruction in Clinical CT. PLoS One, 9(11), (2014).

[19] V. Y. Panin, F. Kehren, C. Michel, and M. Casey. Fully 3-D PET reconstruction with system matrix derived from point source measurements. IEEE Transactions on Medical Imaging, 25(7):907–921, (2006).

[20] E. Y. Sidky, C. Kao, and X. Pan. Accurate image reconstruction from few-views and limited-angle data in divergent-beam CT. Journal of X-Ray Science and Technology, 14(2):119–139, (2006).

[21] X. Jia, B. Dong, Y. Lou, and S. B. Jiang. GPU-based iterative cone-beam CT reconstruction using tight frame regularization. Physics in medicine and biology, 56(13):3787–3807, (2011).

[22] J. I. Agulleiro and J. J. Fernandez. Fast tomographic reconstruction on multicore computers. Bioinformatics, 27(4):582–583, (2011).

[23] X. Li, J. Ni, and G. Wang. Parallel iterative cone beam CT image reconstruction on a PC cluster. Journal of X-Ray Science and Technology, 13(2):63–72, (2005).

[24] J. I. Agulleiro, F. Vázquez, E. M. Garzón, and J. J. Fernández. Hybrid computing: CPU+GPU co-processing and its application to tomographic reconstruction. Ultramicroscopy, 115:109–114, (2012).

[25] T. Goldstein, B. O'Donoghue, S. Setzer, and R. Baraniuk. Fast alternating direction optimization methods. SIAM Journal on Imaging Sciences, 7(3):1588–1623, (2014).

[26] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM journal on imaging sciences, 2(1):183–202, (2009).

[27] R. L. Siddon. Fast calculation of the exact radiological path for a three-dimensional CT array. Medical physics, 12(2):252–255, (1985).

[28] H. Zhao and A. J. Reader. Fast ray-tracing technique to calculate line integral paths in voxel arrays. Nuclear Science Symposium Conference Record, 4:2808–2812, (2003).

[29] J. K. Kim, J. Fessler, Z. Zhang, et al. Forward-projection architecture for fast iterative image reconstruction in X-ray CT. IEEE Transactions on Signal Processing, 60(10):5508–5518, (2012).

[30] A. C. Kak and M. Slaney. Principles of computerized tomographic imaging. Society for Industrial and Applied Mathematics, (2001).

[31] G. N. Hounsfield. A method and apparatus for examination of a body by radiation such as X-ray or gamma radiation, (1972). Patent Specification 1283915.

[32] R. Gordon, R. Bender, and G. T. Herman. Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and X-ray photography. Journal of theoretical Biology, 29(3):471–481, (1970).

[33] Y. L. Mathis, B. Gasharova, and D. Moss. Terahertz radiation at ANKA, the new synchrotron light source in Karlsruhe. Journal of Biological Physics, 29(2):313–318, (2003).

[34] UFO: ultra-fast X-ray imaging. URL http://ufo.kit.edu/index.html.

[35] T. Li, X. Li, J. Wang, J. Wen, H. Lu, J. Hsieh, and Z. Liang. Nonlinear sinogram smoothing for low-dose X-ray CT. IEEE Transactions on Nuclear Science, 51(5): 2505–2513, (2004).

[36] J. Wang, T. Li, H. Lu, and Z. Liang. Penalized weighted least-squares approach to sinogram noise reduction and image reconstruction for low-dose X-ray computed tomography. IEEE Transactions on Medical Imaging, 25(10):1272–1283, (2006).

[37] P. J. La Rivière, J. Bian, and P. A. Vargas. Penalized-likelihood sinogram restoration for computed tomography. IEEE Transactions on Medical Imaging, 25(8): 1022–1036, (2006).

[38] E. Y. Sidky and X. Pan. Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization. Physics in medicine and biology, 53(17):4777–4807, (2008).

[39] R. Marks. Introduction to Shannon sampling and interpolation theory. Springer Science & Business Media, (2012).

[40] X. Yang, T. Jejkal, H. Pasic, R. Stotzka, A. Streit, J. van Wezel, and T. dos Santos Rolo. Data intensive computing of X-ray computed tomography reconstruction at the LSDF. 21st Euromicro International Conference on Parallel, Distributed and Network-Based Processing (PDP), pages 86–93, (2013).

[41] L. A. Shepp and B. F. Logan. The Fourier reconstruction of a head section. IEEE Transactions on Nuclear Science, 21(3):21–43, (1974).

[42] E. J. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. IEEE Transactions on Information Theory, 52(2):489–509, (2006).

[43] T. Chan, S. Esedoglu, F. Park, and A. Yip. Recent developments in total variation image restoration. Mathematical Models of Computer Vision, 17, (2005).

[44] Y. Hu and M. Jacob. Higher degree total variation (HDTV) regularization for image recovery. IEEE Transactions on Image Processing, 21(5):2559–2571, (2012).

[45] M. Avriel. Nonlinear programming: analysis and methods. Dover Publications, Mineola, NY, (2003). ISBN 0486432270.

[46] J. M. Bioucas-Dias and M. A. T. Figueiredo. A new TwIST: two-step iterative shrinkage / thresholding algorithms for image restoration. IEEE Transactions on Image Processing, 16(12):2992–3004, (2007).

[47] S. Becker, J. Bobin, and E.J. Candès. NESTA: a fast and accurate first-order method for sparse recovery. SIAM Journal on Imaging Sciences, 4(1):1–39, (2011).

[48] T. L. Jensen, J. H. Jørgensen, P. C. Hansen, and S. H. Jensen. Implementation of an optimal first-order method for strongly convex total variation regularization. BIT Numerical Mathematics, 52(2):329–356, (2012).

[49] Y. Wang, J. Yang, W. Yin, and Y. Zhang. A new alternating minimization algorithm for total variation image reconstruction. SIAM Journal on Imaging Sciences, 1(3): 248–272, (2008).

[50] C. Li, W. Yin, and Y. Zhang. User's guide for TVAL3: TV minimization by augmented lagrangian and alternating direction algorithms. CAAM Report, (2009).

[51] C. Li. Compressive Sensing for 3D Data Processing Tasks: Applications, Models and Algorithms. PhD Thesis in Rice University, (2011).

[52] J. Yang, Y. Zhang, and W. Yin. A fast alternating direction method for TVL1-L2 signal reconstruction from partial Fourier data. IEEE Journal of Selected Topics in Signal Processing, 4(2):288–297, (2010).

[53] M. Fukushima. Application of the alternating direction method of multipliers to separable convex programming problems. Computational Optimization and Applications, 1(1):93–111, (1992).

[54] T. Goldstein and S. Osher. The split Bregman method for L1-regularized problems. SIAM Journal on Imaging Sciences, 2(2):323–343, (2009).

[55] E. Esser. Applications of Lagrangian-based alternating direction methods and connections to split Bregman. CAM Report, 9:31, (2009).

[56] X. Tai and C. Wu. Augmented Lagrangian method, dual methods and split Bregman iteration for ROF model. Scale Space and Variational Methods in Computer Vision, pages 502–513, (2009).

[57] R. Stotzka, W. Mexner, T. dos Santos Rolo, H. Pasic, J. Van Wezel, V. Hartmann, T. Jejkal, A. Garcia, D. Haas, and A. Streit. Large scale data facility for data intensive synchrotron beamlines. Proceedings of ICALEPCS2011, pages 1216–1219, (2011).

[58] A. O. Garcia, S. Bourov, A. Hammad, J. van Wezel, B. Neumair, A. Streit, V. Hartmann, T. Jejkal, P. Neuberger, and R. Stotzka. The large scale data facility: Data intensive computing for scientific experiments. IEEE International Symposium on Parallel and Distributed Processing Workshops and Phd Forum (IPDPSW), pages 1467–1474, (2011).

[59] P. C. Hansen and D. P. O'Leary. The use of the L-curve in the regularization of discrete ill-posed problems. SIAM Journal on Scientific Computing, 14(6):1487–1503, (1993).

[60] V. A. Morozov. On the solution of functional equations by the method of regularization. Soviet Mathematics Doklady, 7(1):414–417, (1966).

[61] G. H. Golub, M. Heath, and G. Wahba. Generalized cross-validation as a method for choosing a good ridge parameter. Technometrics, 21(2):215–223, (1979).

[62] L. Reichel and G. Rodriguez. Old and new parameter choice rules for discrete ill-posed problems. Numerical Algorithms, 63(1):65–87, (2013).

[63] A. Sakellariou, T. J. Sawkins, T. J. Senden, and A. Limaye. X-ray tomography for mesoscale physics applications. Physica A: Statistical Mechanics and its Applications, 339(1):152–158, (2004).

[64] T. M. Buzug. Computed tomography: from photon statistics to modern cone-beam CT. Springer Science & Business Media, (2008).

[65] J. Fessler. Analytical tomographic image reconstruction methods. Image Reconstruction: Algorithms and Analysis, pages 1–47, (2009).

[66] W. Zhuang, S. S. Gopal, and T. J. Hebert. Numerical evaluation of methods for computing tomographic projections. IEEE Transactions on Nuclear Science, 41 (4):1660–1665, (1994).

[67] R. Li, J. Pan, and Z. Hui. Accurate Ray-driven Algorithm in Iterative Image Reconstruction for Cone-beam CT. Journal of Computational Information Systems, 7(13):4686–4693, (2011).

[68] B. De Man and S. Basu. Distance-driven projection and backprojection in three dimensions. Physics in Medicine and Biology, 49(11):2463, (2004).

[69] S. Basu and B. De Man. Branchless distance driven projection and backprojection. Electronic Imaging, pages 60650Y–60650Y, (2006).

[70] G. L. Zeng and G. T. Gullberg. A ray-driven backprojector for backprojection filtering and filtered backprojection algorithms. Nuclear Science Symposium and Medical Imaging Conference, pages 1199–1201, (1993).

[71] B. De Man and S. Basu. Distance-driven projection and backprojection. Nuclear Science Symposium Conference Record, 3:1477–1480, (2002).

[72] G. N. Hounsfield. Computerized transverse axial scanning (tomography): Part 1. Description of system. The British Journal of Radiology, 46(552):1016–1022, (1973).

[73] L. A. Feldkamp, L. C. Davis, and J. W. Kress. Practical cone-beam algorithm. Journal of the Optical Society of America, 1(6):612–619, (1984).

[74] A. H. Andersen and A. C. Kak. Simultaneous algebraic reconstruction technique (SART): a superior implementation of the ART algorithm. Ultrasonic imaging, 6 (1):81–94, (1984).

[75] B. E. Oppenheim. Reconstruction tomography from incomplete projections. Reconstruction tomography in diagnostic radiology and nuclear medicine, (1977).

[76] P. Gilbert. Iterative methods for the three-dimensional reconstruction of an object from projections. Journal of theoretical biology, 36(1):105–117, (1972).

[77] A. J. Rockmore and A. Macovski. A maximum likelihood approach to emission image reconstruction from projections. IEEE Transactions on Nuclear Science, 23 (4):1428–1432, (1976).

[78] M. Jiang and G. Wang. Development of iterative algorithms for image reconstruction. Journal of X-ray Science and Technology, 10(1):77–86, (2001).

[79] A. P. Dempster, N. M. Lamb, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society, 39(1): 1–38, (1977).

[80] H. M. Hudson and R. S. Larkin. Accelerated image reconstruction using ordered subsets of projection data. IEEE Transactions on Medical Imaging, 13(4):601–609, (1994).

[81] S. H. Manglos, G. M. Gagne, A. Krol, F. D. Thomas, and R. Narayanaswamy. Transmission maximum-likelihood reconstruction with ordered subsets for cone beam CT. Physics in Medicine and Biology, 40(7):1225, (1995).

[82] L. A. Shepp and Y. Vardi. Maximum likelihood reconstruction for emission tomography. IEEE Transactions on Medical Imaging, 1(2):113–122, (1982).

[83] S. Kaczmarz. Angenäherte Auflösung von Systemen Linearer Gleichungen. Bulletin International de l'Academie Polonaise des Sciences et des Lettres, 35:355–357, (1937).

[84] S. Kaczmarz. Approximate solution of systems of linear equations. International Journal of Control, 57(6):1269–1271, (1993).

[85] K. Lange, R. Carson, et al. EM reconstruction algorithms for emission and transmission tomography. Journal of Computer Assisted Tomography, 8(2):306–316, (1984).

[86] C. Chen, C. E. Metz, and X. Hu. Maximum likelihood reconstruction in PET and TOFPET. Mathematics and Computer Science in Medical Imaging, pages 319–329, (1988).

[87] M. I. Miller, D. L. Snyder, and T. R. Miller. Maximum-Likelihood Reconstruction for Single-Photon Emission Computed-Tomography. IEEE Transactions on Nuclear Science, 32(1):769–778, (1985).

[88] T. J. Holmes. Maximum-likelihood image restoration adapted for noncoherent optical imaging. Journal of the Optical Society of America, 5(5):666–673, (1988).

[89] J. A. Browne and T. J. Holmes. Developments with maximum likelihood X-ray computed tomography. IEEE Transactions on Medical Imaging, 11(1):40–52, (1992).

[90] Q. Xu, H. Yu, X. Mou, L. Zhang, J. Hsieh, and G. Wang. Low-dose X-ray CT reconstruction via dictionary learning. IEEE Transactions on Medical Imaging, 31 (9):1682–1697, (2012).

[91] H. Yan, L. Cervino, X. Jia, and S. B. Jiang. A comprehensive study on the relationship between the image quality and imaging dose in low-dose cone beam CT. Physics in Medicine and Biology, 57(7):2063, (2012).

[92] B. Dong, J. Li, and Z. Shen. X-ray CT image reconstruction via wavelet frame based regularization and Radon domain inpainting. Journal of Scientific Computing, 54(2-3):333–349, (2013).

[93] C. R. Vogel and M. E. Oman. Iterative methods for total variation denoising. SIAM Journal on Scientific Computing, 17(1):227–238, (1996).

[94] Y. You and M. Kaveh. A regularization approach to joint blur identification and image restoration. IEEE Transactions on Image Processing, 5(3):416–428, (1996).

[95] R. Dapp. Abbildungsmethoden für die Brust mit einem 3D-Ultraschall-Computertomographen. PhD thesis, Karlsruher Institut für Technologie (KIT), (2013).

[96] A. Hormati, I. Jovanovic, O. Roy, and M. Vetterli. Robust ultrasound travel-time tomography using the bent ray model. SPIE Medical Imaging, 7629:76290I, (2010).

[97] R. Battiti. First-and second-order methods for learning: between steepest descent and Newton's method. Neural computation, 4(2):141–166, (1992).

[98] W. W. Hager and H. Zhang. A survey of nonlinear conjugate gradient methods. Pacific journal of Optimization, 2(1):35–58, (2006).

[99] I. Daubechies, M. Defrise, and C. De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. Communications on Pure and Applied Mathematics, 57(11):1413–1457, (2004).

[100] M. A. T. Figueiredo and R. D. Nowak. An EM algorithm for wavelet-based image restoration. IEEE Transactions on Image Processing, 12(8):906–916, (2003).

[101] C. Vonesch and M. Unser. A fast iterative thresholding algorithm for wavelet-regularized deconvolution. Optical Engineering + Applications, page 67010D, (2007).

[102] D. Stsepankou, A. Arns, S. K. Ng, P. Zygmanski, and J. Hesser. Evaluation of robustness of maximum likelihood cone-beam CT reconstruction with total variation regularization. Physics in Medicine and Biology, 57(19):5955, (2012).

[103] R. P. V. Rao. Parallel implementation of the filtered back projection algorithm for tomographic imaging. Virginia Tech, (1995).

[104] J. Rajruangrabin, S. S. Thongvigitmanee, S. Aootaphao, and P. Thajchayapong. Acceleration of filtered back-projection algorithm for 3D cone-beam CT reconstruction using parallel computation. Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), pages 1–4, (2013).

[105] S. Frey, C. Müller, M. Strengert, and T. Ertl. Concurrent CT reconstruction and visual analysis using hybrid multi-resolution raycasting in a cluster environment. Advances in Visual Computing, pages 357–366, (2009).

[106] E. J. Candès and M. B. Wakin. An introduction to compressive sampling. IEEE Signal Processing Magazine, 25(2):21–30, (2008).

[107] D. Liang, H. Wang, Y. Chang, and L. Ying. Sensitivity encoding reconstruction with nonlocal total variation regularization. Magnetic resonance in medicine, 65 (5):1384–1392, (2011).

[108] Y. Liu, J. Ma, Y. Fan, and Z. Liang. Adaptive-weighted total variation minimization for sparse data toward low-dose x-ray computed tomography image reconstruction. Physics in Medicine and Biology, 57(23):7923, (2012).

[109] M. Lustig, D. Donoho, and J. M. Pauly. Sparse MRI: the application of compressed sensing for rapid MR imaging. Magnetic Resonance in Medicine, 58(6):1182–1195, (2007).

[110] X. Yang, R. Hofmann, R. Dapp, T. van de Kamp, T. dos Santos Rolo, X. Xiao, J. Moosmann, J. Kashef, and R. Stotzka. TV-based conjugate gradient method and discrete L-curve for few-view CT reconstruction of X-ray in vivo data. Optics Express, 23(5):5368–5387, (2015).

[111] Y. Dai and Y. Yuan. A nonlinear conjugate gradient method with a strong global convergence property. SIAM Journal on Optimization, 10(1):177–182, (1999).

[112] C. Kübel, D. Niemeyer, R. Cieslinski, and S. Rozeveld. Electron tomography of nanostructured materials–towards a quantitative 3D analysis with nanometer resolution. Materials Science Forum, 638:2517–2522, (2010).

[113] X. Yang, J. Moosmann, J. Kashef, R. Stotzka, and R. Hofmann. Application of sparse CT reconstruction to developmental biology. IEEE International Symposium on Biomedical Imaging (ISBI), (2014).

[114] X. Yang, T. Jejkal, R. Stotzka, Pasic H., T. dos Santos Rolo, T. van de Kamp, and A. Streit. Algebraic reconstruction of ultrafast tomography images at the large scale data facility. 14th International conference on accelerator & large experimental physics control systems (ICALEPCS2013), (2013).

[115] Y. Lin. Application of the UPRE method to optimal parameter selection for large scale regularization problems. IEEE Southwest Symposium on Image Analysis and Interpretation (SSIAI), pages 89–92, (2008).

[116] C. R. Vogel. Non-convergence of the L-curve regularization parameter selection method. Inverse Problems, 12(4):535, (1996).

[117] D. Calvetti, S. Morigi, L. Reichel, and F. Sgallari. Tikhonov regularization and the L-curve for large discrete ill-posed problems. Journal of Computational and Applied Mathematics, 123(1):423–446, (2000).

[118] R. Hofmann, A. Schober, J. Moosmann, M. Hertel, S. Hahn, V. Weinhardt, D. Hänschke, L. Helfen, X. Xiao, and T. Baumbach. Single-distance livecell imaging with propagation-based X-ray phase contrast. (to be submitted).

[119] The energy dependence of the complex refractive index can be obtained for a variety of chemical compounds. URL http://henke.lbl.gov/optical_constants/getdb2.html.

[120] Z. Wang and A. C. Bovik. Mean squared error: love it or leave it? A new look at signal fidelity measures. IEEE Signal Processing Magazine, 26(1):98–117, (2009).

[121] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing, 13(4):600–612, (2004).

[122] R. Liu. Data compression in ultrasound computed tomography. PhD thesis, Karlsruher Institut für Technologie (KIT), (2011).

[123] S. Chilingaryan, A. Kopmann, A. Mirone, and T. dos Santos Rolo. A GPU–based architecture for real-time data assessment at synchrotron experiments. Real Time Conference (RT2010), pages 1–8, (2010).

[124] F. Marone, C. Hintermuller, R. Geus, and M. Stampanoni. Towards real-time tomography: fast reconstruction algorithms and GPU implementation. Nuclear Science Symposium Conference Record, pages 555–561, (2008).

[125] Y. Okitsu, F. Ino, and K. Hagihara. High–performance cone beam reconstruction using CUDA compatible GPUs. Parallel Computing, 36(2-3):129–141, (2010).

[126] J. Treibig, G. Hager, H. G. Hofmann, J. Hornegger, and G. Wellein. Pushing the limits for medical image reconstruction on recent standard multicore processors. International Journal of High Performance Computing Applications, 27(2):162–177, (2013).

[127] J. Ni, X. Li, T. He, and G. Wang. Review of Parallel Computing Techniques for Computed Tomography Image Reconstruction. Current Medical Imaging Reviews, 2(4):405–414, (2006).

[128] T. Jejkal, V. Hartmann, R. Stotzka, J. Otte, A. García, J. van Wezel, and A. Streit. LAMBDA–the LSDF execution framework for data intensive applications. 20th Euromicro International Conference on Parallel, Distributed and Network–Based Processing (PDP), pages 213–220, (2012).

[129] KIT Data Manager. URL http://ipelsdf1.lsdf.kit.edu/cms/index.php/nav-pro-projects/nav-pro-act-kdm.

[130] X. Pan, E. Y. Sidky, and M. Vannier. Why do commercial CT scanners still employ traditional, filtered back-projection for image reconstruction? Inverse problems, 25(12):123009, (2009).

[131] Y. Nesterov. Gradient methods for minimizing composite objective function. Technical report, Université catholique de Louvain, Center for Operations Research and Econometrics (CORE), (2007).

[132] Z. Tian, X. Jia, K. Yuan, T. Pan, and S. B. Jiang. Low-dose CT reconstruction via edge-preserving total variation regularization. Physics in Medicine and Biology, 56(18):5949, (2011).

[133] Y. Dong, M. Hintermüller, and M. M. Rincon-Camacho. Automated regularization parameter selection in multi-scale total variation models for image restoration. Journal of Mathematical Imaging and Vision, 40(1):82–104, (2011).

[134] The Apache Software Foundation. Welcome to Apache Hadoop!, (March 2012). URL http://hadoop.apache.org/.

[135] J. Dean and S. Ghemawat. MapReduce: simplified data processing on large clusters. Communications of the ACM, 51(1):107–113, (2008).

[136] FUSE: Filesystem in Userspace. URL http://fuse.sourceforge.net/.

[137] IBM General Parallel File System. URL http://www.ibm.com/systems/software/gpfs/.

[138] Scientific Linux. URL http://www.scientificlinux.org/.

[139] L.F. John and W.O. Peter. Modules – Software Environment Management, (October 2010). URL http://www.ibm.com/systems/software/gpfs/.

List of abbreviations

ADMM	Alternating direction method of multipliers
ADS	Adaptive step-size
AIRT	Simultaneous Iterative Reconstruction Technique
ANKA	Synchrotron Radiation Facility at KIT
APS	Advanced Photon Source
ART	Algebraic Reconstruction Technique
AwTV	Adaptive-weighted Total Variation
CG	Conjugate Gradient
CGTV	Conjugate gradient-based reconstruction algorithm with TV regularization
CPU	Central Processing Unit
CS	Compressive Sampling/Sensing
CT	Computed Tomography
DAQ	Data Acquisition
DFT	Discrete Fourier Transform
DGT	Discrete Gradient Transform
DP	Discrepancy Principle
DWT	Discrete Wavelet Transform
EM	Expectation Maximization
ET	Electron Tomography
FBP	Filtered Back-projection
FDK	Feldkamp, Davis and Kress
FISTA	Fast iterative shrinkage-thresholding algorithm
FT	Fourier Transform
GCV	Generalized Cross Validation
GPU	Graphics Processor Unit
IFT	Inverse Fourier Transform
IQA	Image Quality Assessment
IR	Iterative Reconstruction
IST	Iterative shrinkage/thresholding algorithm
KIT	Karlsruhe Institute of Technology

LSDF	Large Scale Data Facility
ML	Maximum Likelihood
MRI	Magnetic Resonance Imaging
MSE	Mean Square Error
NESTA	Nesterov's algorithm
NLTV	Nonlocal Total Variation
OI	Optical Imaging
OS	ordered Subset
PET	Positron Emission Tomography
POCS	Projection on convex sets
RecRF	Reconstruction from partial Fourier data
RIP	Restricted Isometry property
SART	Simultaneous Algebraic Reconstruction Technique
SEM	Scanning Electron Microscope
SNR	Signal to Noise Ratio
SPECT	Single Photon Emission Computed Tomography
SSIM	Structural Similarity
SVD	Single Value Decomposition
TIF	Tagged Image File Format
TOPO-TOMO	Topography and Tomography
TV	Total Variation
TVAL3	TV minimization scheme based on augmented Lagrangian and alternating direction algorithms
TwIST	Two-step iterative shrinkage-thresholding algorithm
UFO	Ultra-fast X-ray imaging
UPN	Unknown Parameter Nesterov
UPRE	Unbiased Predictive Risk Estimator
XPμCT	X-ray Phase-contrast Computed Microtomography

List of symbols

Forward problem in computed tomography (continuous)

Δx	Small distance in x-axis
μ, v	μ- and v-axis of Cartesian system in Fourier domain
F	Fourier transform of the object function f
P	Fourier transform of the projection function p
ω, θ	Radial and angular coordinates of polar system in Fourier domain
u	X-ray attenuation index in one dimension
I_0	Input photon intensity
l	straight line along which the X-ray is propagating
f	Continuous object function representing the distribution of attenuation index
I	output photon intensity
p	Projection function
r	Radial coordinate in polar system
$\Re$	Radon transform
θ	Angle in which projections are measured
x, y, z	x-, y-, and z-axis

Computed tomography reconstruction (discrete)

Φ	Base changes in a sparse domain
Ψ	Basis used for sensing/measuring image $\mathbf{x}$
C	Number of all image elements in $\mathbf{x}$ or column number of system matrix $\mathbf{A}$
N	Column number of the image $\mathbf{X}$
$F(\mathbf{x})$	Data fidelity term with respect to image $\mathbf{x}$ in minimization program

a_{ij}	Element of system matrix $\mathbf{A}$ representing the contribution of the i_{th} cell x_i in $\mathbf{x}$ to the j_{th} projection element p_j in $\mathbf{p}$
b_i	i_{th} element of $\mathbf{b}$
ε	Error vector in projections
Γ	Shrinkage/soft-thresholding operator
x_i	i_{th} element in image $\mathbf{x}$
X	Elements of image $\mathbf{X}$
$\mathbf{X}$	Matrix formulation of the image representing attenuation coefficients of the object
$\mathbf{x}$	Vector formulation of the image representing attenuation coefficients of the object
λ	Trade-off parameter
$\mathbf{A}$	Transform matrix of a CT system
S	Size of each projection
P	Number of projections
$f(\mathbf{x})$	Objective function with respect to image $\mathbf{x}$ in minimization program
$\mathbf{x}^*$	Optimal image of $\mathbf{x}$
λ^*	Optimal value of λ
p_j	j_{th} element in projection $\mathbf{p}$
$\mathbf{p}$	Vector formulation of measured projections
$R(\mathbf{x})$	Regularization function with respect to image $\mathbf{x}$
R	Number of all projection elements in $\mathbf{p}$ or row number of system matrix $\mathbf{A}$
M	Row number of the image $\mathbf{X}$
K	Size of each projection
$\mathbf{x}'$	Sparse representation of $\mathbf{x}$
ϵ	Tolerance parameter
$\|\mathbf{x}\|_{\text{TV}}$	Total variation of image $\mathbf{x}$
$T(x)$	Total variation term with respect to image $\mathbf{x}$ in minimization program
$\mathbf{b}$	An arbitrary vector
L	Linear dimension of the field of view

Publications related to this thesis

1. X. Yang, R. Hofmann, R. Dapp, T. van de Kamp, T. dos Santos Rolo, X. Xiao, J. Moosmann, J. Kashef, and R. Stotzka. TV-based Conjugate Gradient Method and Discrete L-curve for Few-view CT Reconstruction of X-ray *in vivo* Data, Optic Express, Vol. 23, Issue 5, pp.5368-5387, (2015)
   DOI: 10.1364/OE.23.005368

2. X. Yang, J. Moosmann, J. Kashef, R. Stotzka, and R. Hofmann. Application of Sparse CT Reconstruction to Developmental Biology, IEEE International Symposium on Biomedical Imaging (ISBI), (2014)

3. X. Yang, T. Jejkal, H. Pasic, R. Stotzka, A. Streit, J. van Wezel, T. dos Santos Rolo. Data Intensive Computing of X-ray Computed Tomography Reconstruction at the LSDF, 21st Euromicro International Conference on Parallel, Distributed, and Network-Based Processing (PDP), IEEE, pp.86-93, (2013)
   DOI: 10.1109/PDP.2013.21

4. X. Yang, T. Jejkal, R. Stotzka, H. Pasic, T. dos Santos Rolo, T. van de Kamp, A. Streit. Algebraic Reconstruction of Ultrafast Tomography Images at the Large Scale Data Facility, Proceedings of the 14th International Conference on Accelerator & Large Experimental Physics Control Systems (ICALEPCS), (2013)

Acknowledgments

First of all I would like to express my gratitude to my advisor Dr. Rainer Stotzka for providing me an opportunity to join the Institute for Data Processing and Electronics (IPE) at the Karlsruhe Institute of Technology (KIT) as a PhD student, for his professional and moral support with respect to my PhD research. I am very grateful for his continuous encouragement and assistance in my study and even in my new life in Germany during the last four years of my PhD period.

Also many thanks to my supervising professor Prof. Marc Weber for his support in my PhD study, for his insightful comments and encouragement.

My sincere thanks also go to my colleagues in the Software Method (SWM) group: Robin Happ, Thomas Jejkal, Halil Pasic, Volker Hartmann, Francesca Rindone, Danah Tonne, Swati Chandna, and Ajinkya Prabhune. Their ideas inspired me a lot in my daily work and I indeed learned a lot from the nice discussion with them. I am thankful for their suggestions and encouragements. Special thanks to Francesca Rindone who additionally helped me to get used to the life in Germany.

This work would not have been possible without the data contributed by our partners at the Institute for Photon Science and Synchrotron Radiation (IPS) and the Institute of Nanotechnology. Many thanks to Dr. Ralf Hofmann, Dr. Thomas van de Kamp, Dr. Christian Kübel, Tomy dos Santos Rolo, and Julian Mossmann for providing me a large amount of interesting tomographic data.

Finally, I would like to thank my family and friends. Special thanks to my husband for supporting me spiritually throughout my PhD study and my life in general. Also thanks to my parents and parents-in-law for their understanding and supporting.